

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
7 April 2005 (07.04.2005)

PCT

(10) International Publication Number
WO 2005/030039 A2

(51) International Patent Classification⁷: **A61B**
(21) International Application Number:
PCT/US2004/031481
(22) International Filing Date:
23 September 2004 (23.09.2004)
(25) Filing Language: English
(26) Publication Language: English
(30) Priority Data:
60/505,527 23 September 2003 (23.09.2003) US

(71) Applicant (for all designated States except US): UNI-
VERSITY OF NORTH CAROLINA AT CHAPEL
HILL [US/US]; 308 Bynum Hall, Campus Box 4105,
Chapel Hill, North Carolina 27599-4105 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): STAFFORD, Dar-
rel, W. [US/US]; 300 Rainbow Drive, Carrboro, North
Carolina 27510 (US). LI, Tao [CN/US]; 7585 Charmant
Drive, Apt. 807, San Diego, California 92122 (US).

(74) Agent: MYERS BIGEL SIBLEY & SAJOVEC, P.A.?;
P.O. Box 37428, Raleigh, North Carolina 27627 (US).

(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN,
CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI,
GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE,
KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD,
MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG,
PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM,
TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM,
ZW.

(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM,
ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,
FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI,
SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished
upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

(54) Title: METHODS AND COMPOSITIONS FOR THE CORRELATION OF SINGLE NUCLEOTIDE POLYMORPHISMS IN
THE VITAMIN K EPOXIDE REDUCTASE GENE AND WARFARIN DOSAGE

(57) Abstract: The present invention provides a method of identifying a human subject having increased or decreased sensitivity
to warfarin, comprising detecting in the subject the presence of a single nucleotide polymorphism in the VKOR gene, wherein the
single nucleotide polymorphism is correlated with increased or decreased sensitivity to warfarin, thereby identifying the subject
having increased or decreased sensitivity to warfarin.

WO 2005/030039 A2

**METHODS AND COMPOSITIONS FOR THE CORRELATION OF SINGLE
NUCLEOTIDE POLYMORPHISMS IN THE VITAMIN K EPOXIDE REDUCTASE
GENE AND WARFARIN DOSAGE**

5

Statement of Priority

The present application claims the benefit, under 35 U.S.C. § 119(e), of U.S. Provisional Application Serial No. 60/505,527, filed September 23, 2003, the entire contents of which are incorporated by reference herein.

10

Government Support

The present invention was made, in part, with the support of grant numbers 5P01 HL06350-42 and 5-R01 HL48318 from the National Institutes of Health. The United States Government has certain rights to this invention.

15

Field of the Invention

The present invention concerns isolated nucleic acids, host cells containing the same, and methods of use thereof, as well as methods and compositions directed to identification of single nucleotide polymorphisms (SNPs) in the Vitamin K epoxide reductase (VKOR) gene and their correlation with sensitivity to warfarin.

20

Background of the Invention

The function of numerous proteins requires the modification of multiple glutamic acid residues to γ -carboxyglutamate. Among these vitamin K-dependent (VKD) coagulation proteins, FIX (Christmas factor), FVII, and prothrombin are the best known. The observation that a knock-out of the gene for matrix Gla protein results in calcification of the mouse's arteries (Luo et al. (1997) "Spontaneous calcification of arteries and cartilage in mice lacking matrix GLA protein" *Nature* 386:78-81) emphasizes the importance of the vitamin K cycle for proteins with functions other than coagulation. Moreover, Gas6 and other Gla proteins of unknown function are expressed in neural tissue and warfarin exposure *in utero* results in mental retardation and facial abnormalities. This is consistent with the observation that the expression of VKD carboxylase, the enzyme that accomplishes the Gla modification, is temporally regulated in a tissue-specific manner with high

25

30

expression in the nervous system during early embryonic stages. Concomitant with carboxylation, reduced vitamin K, a co-substrate of the reaction, is converted to vitamin K epoxide. Because the amount of vitamin K in the human diet is limited, vitamin K epoxide must be converted back to vitamin K by vitamin K epoxide reductase (VKOR) to prevent its depletion. Warfarin, the most widely used anticoagulation drug, targets VKOR and prevents the regeneration of vitamin K. The consequence is a decrease in the concentration of reduced vitamin K, which results in a reduced rate of carboxylation by the γ -glutamyl carboxylase and in the production of undercarboxylated vitamin K-dependent proteins.

In the United States alone, warfarin is prescribed to more than one million patients per year and in Holland, it has been reported that approximately 2% of the population is on long term warfarin therapy. Because the dose of warfarin required for a therapeutic level of anticoagulation varies greatly between patients, the utilization of warfarin is accompanied by a significant risk of side effects. For example, it has been reported that following initiation of warfarin therapy, major bleeding episodes occurred in 1-2% of patients and death occurred in 0.1-0.7 % of patients. In spite of the dangers, it has been estimated that warfarin use can prevent 20 strokes per induced bleeding episode and is probably underutilized because of the fear of induced bleeding.

The present invention overcomes previous shortcomings in the art by providing methods and compositions for correlating single nucleotide polymorphisms in a subject with an increased or decreased sensitivity to warfarin, thereby allowing for more accurate and rapid determination of therapeutic and maintenance doses of warfarin at reduced risk to the subject.

Summary of the Invention

The present invention provides a method of identifying a human subject having increased or decreased sensitivity to warfarin, comprising detecting in the subject the presence of a single nucleotide polymorphism in the VKOR gene, wherein the single nucleotide polymorphism is correlated with increased or decreased sensitivity to warfarin, thereby identifying the subject having increased or decreased sensitivity to warfarin.

Additionally provided is a method of identifying a human subject having increased or decreased sensitivity to warfarin, comprising: a) correlating the

presence of a single nucleotide polymorphism in the VKOR gene with increased or decreased sensitivity to warfarin; and b) detecting the single nucleotide polymorphism of step (a) in the subject, thereby identifying a subject having increased or decreased sensitivity to warfarin.

5 In a further embodiment, the present invention provides a method of identifying a single nucleotide polymorphism in the VKOR gene correlated with increased or decreased sensitivity to warfarin, comprising:

a) identifying a subject having increased or decreased sensitivity to warfarin;

10 b) detecting in the subject the presence of a single nucleotide polymorphism in the VKOR gene; and

c) correlating the presence of the single nucleotide polymorphism of step (b) with the increased or decreased sensitivity to warfarin in the subject, thereby identifying a single nucleotide polymorphism in the VKOR gene correlated with increased or decreased sensitivity to warfarin.

15 In addition, the present invention provides a method of correlating a single nucleotide polymorphism in the VKOR gene of a subject with increased or decreased sensitivity to warfarin, comprising: a) identifying a subject having increased or decreased sensitivity to warfarin; b) determining the nucleotide sequence of the VKOR gene of the subject of (a); c) comparing the nucleotide sequence of step (b)
20 with the wild type nucleotide sequence of the VKOR gene; d) detecting a single nucleotide polymorphism in the nucleotide sequence of (b); and e) correlating the single nucleotide polymorphism of (d) with increased or decreased sensitivity to warfarin in the subject of (a).

A further aspect of the present invention is an isolated nucleic acid encoding
25 vitamin K epoxide reductase (VKOR), particularly mammalian (e.g., human, ovine, bovine, monkey, etc.) VKOR. Examples include (a) nucleic acids as disclosed herein, such as isolated nucleic acids having the nucleotide sequence as set forth in SEQ ID NO: 8 or SEQ ID NO: 9; (b) nucleic acids that hybridize to isolated nucleic acids of (a) above or the complement thereof (e.g., under stringent conditions),
30 and/or have substantial sequence identity to nucleic acids of (a) above (e.g., are 80, 85, 90 95 or 99% identical to nucleic acids of (a) above), and encode a VKOR; and (c) nucleic acids that differ from the nucleic acids of (a) or (b) above due to the degeneracy of the genetic code, but code for a VKOR encoded by a nucleic acid of (a) or (b) above.

The term "stringent" as used here refers to hybridization conditions that are commonly understood in the art to define the commodities of the hybridization procedure. Stringency conditions can be low, high or medium, as those terms are commonly know in the art and well recognized by one of ordinary skill. High stringency hybridization conditions that will permit homologous nucleotide sequences to hybridize to a nucleotide sequence as given herein are well known in the art. As one example, hybridization of such sequences to the nucleic acid molecules disclosed herein can be carried out in 25% formamide, 5X SSC, 5X Denhardt's solution and 5% dextran sulfate at 42°C, with wash conditions of 25% formamide, 5X SSC and 0.1% SDS at 42°C, to allow hybridization of sequences of about 60% homology. Another example includes hybridization conditions of 6X SSC, 0.1% SDS at about 45°C, followed by wash conditions of 0.2X SSC, 0.1% SDS at 50-65°C. Another example of stringent conditions is represented by a wash stringency of 0.3 M NaCl, 0.03M sodium citrate, 0.1% SDS at 60-70°C using a standard hybridization assay (see SAMBROOK et al., EDS., MOLECULAR CLONING: A LABORATORY MANUAL 2d ed. (Cold Spring Harbor, NY 1989, the entire contents of which are incorporated by reference herein). In various embodiments, stringent conditions can include, for example, highly stringent (i.e., high stringency) conditions (e.g., hybridization to filter-bound DNA in 0.5 M NaHPO₄, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C., and washing in 0.1xSSC/0.1% SDS at 68°C.), and/or moderately stringent (i.e., medium stringency) conditions (e.g., washing in 0.2xSSC/0.1% SDS at 42°C.).

An additional aspect of the present invention is a recombinant nucleic acid comprising a nucleic acid encoding vitamin K epoxide reductase as described herein operatively associated with a heterologous promoter.

A further aspect of the present invention is a cell that contains and expresses a recombinant nucleic acid as described above. Suitable cells include plant, animal, mammal, insect, yeast and bacterial cells.

A further aspect of the present invention is an oligonucleotide that hybridizes to an isolated nucleic acid encoding VKOR as described herein.

A further aspect of the present invention is isolated and purified VKOR (e.g., VKOR purified to homogeneity) encoded by a nucleic acid as described herein. For example, the VKOR of this invention can comprise the amino acid sequence as set forth in SEQ ID NO:10.

A further aspect of the present invention is a method of making a vitamin K dependent protein which comprises culturing a host cell that expresses a nucleic acid encoding a vitamin K dependent protein in the presence of vitamin K and produces a vitamin K dependent protein, and then harvesting the vitamin K dependent protein from the culture, the host cell containing and expressing a heterologous nucleic acid encoding vitamin K dependent carboxylase, and the host cell further containing and expressing a heterologous nucleic acid encoding vitamin K epoxide reductase (VKOR) and producing VKOR as described herein.

Brief Description of the Drawings

Figures 1A-D Comparisons of warfarin dosages in wild type, heterozygous and homozygous subjects for SNPs vk 2581, vk3294 and vk4769, as well as a comparison of warfarin dosages in wild type and heterozygous subjects for P450 2Y9.

Figure 2. For each of the 13 siRNA pools, three T7 flasks containing A549 cells were transfected and VKOR activity determined after 72 h. The VKOR assay used 25 μ M vitamin K epoxide. One siRNA pool specific for gene gi:13124769 reduced VKOR activity by 64%-70% in eight repetitions.

Figure 3. Time course of inhibition of VKOR activity by the siRNA pool specific for gi:13124769 in A549 cells. VKOR activity decreased continuously during this time period while the level of its mRNA decreased rapidly to about 20% of normal. 25 μ M vitamin K epoxide was used for this assay. The siRNA did not affect the activity of VKD carboxylase or the level of lamin A/C mRNA.

Figure 4. VKOR activity was detected when mGC_11276 was expressed in Sf9 insect cells. $\sim 1 \times 10^6$ cells were used in this assay. Reactions were performed using 32 μ M KO at 30°C for 30 minutes in Buffer D. Blank Sf9 cells served as a negative control and A549 cells as a reference.

Figure 5. Inhibition of VKOR by warfarin. Reactions were performed using 1.6 mg microsomal proteins made from VKOR_Sf9 cells, 60 μ M KO, and various concentration of warfarin at 30°C for 15 minutes in Buffer D.

Detailed Description of the Invention

As used herein, "a," "an" or "the" can mean one or more than one. For example, "a" cell can mean a single cell or a multiplicity of cells.

The present invention is explained in greater detail below. This description is not intended to be a detailed catalog of all the different ways in which the invention may be implemented, or all the features that may be added to the instant invention. For example, features illustrated with respect to one embodiment may be
5 incorporated into other embodiments, and features illustrated with respect to a particular embodiment may be deleted from that embodiment. In addition, numerous variations and additions to the various embodiments suggested herein will be apparent to those skilled in the art in light of the instant disclosure which do not depart from the instant invention. Hence, the following specification is intended to
10 illustrate some particular embodiments of the invention, and not to exhaustively specify all permutations, combinations and variations thereof.

The "Sequence Listing" attached hereto forms a part of the instant specification as if fully set forth herein.

The present invention may be carried out based on the instant disclosure and
15 further utilizing methods, components and features known in the art, including but not limited to those described in US Patent No. 5,268,275 to Stafford and Wu and US Patent No. 6,531,298 to Stafford and Chang, the disclosures of which are incorporated by reference herein in their entirety as if fully set forth herein.

As used herein, "nucleic acids" encompass both RNA and DNA, including
20 cDNA, genomic DNA, synthetic (e.g., chemically synthesized) DNA and chimeras of RNA and DNA. The nucleic acid may be double-stranded or single-stranded. Where single-stranded, the nucleic acid may be a sense strand or an antisense strand. The nucleic acid may be synthesized using oligonucleotide analogs or derivatives (e.g., inosine or phosphorothioate nucleotides). Such oligonucleotides can be used, for
25 example, to prepare nucleic acids that have altered base-pairing abilities or increased resistance to nucleases.

An "isolated nucleic acid" is a DNA or RNA that is not immediately contiguous with both of the coding sequences with which it is immediately contiguous (one on the 5' end and one on the 3' end) in the naturally occurring genome of the organism
30 from which it is derived. Thus, in one embodiment, an isolated nucleic acid includes some or all of the 5' non-coding (e.g., promoter) sequences that are immediately contiguous to the coding sequence. The term therefore includes, for example, a recombinant DNA that is incorporated into a vector, into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote, or which

exists as a separate molecule (e.g., a cDNA or a genomic DNA fragment produced by PCR or restriction endonuclease treatment), independent of other sequences. It also includes a recombinant DNA that is part of a hybrid gene encoding an additional polypeptide sequence.

- 5 The term "isolated" can refer to a nucleic acid or polypeptide that is substantially free of cellular material, viral material, or culture medium (when produced by recombinant DNA techniques), or chemical precursors or other chemicals (when chemically synthesized). Moreover, an "isolated nucleic acid fragment" is a nucleic acid fragment that is not naturally occurring as a fragment and
10 would not be found in the natural state.

The term "oligonucleotide" refers to a nucleic acid sequence of at least about six nucleotides to about 100 nucleotides, for example, about 15 to 30 nucleotides, or about 20 to 25 nucleotides, which can be used, for example, as a primer in a PCR amplification or as a probe in a hybridization assay or in a microarray.

- 15 Oligonucleotides may be natural or synthetic, e.g., DNA, RNA, modified backbones, etc.

Where a particular nucleotide sequence is said to have a specific percent identity to a reference nucleotide sequence, the percent identity is relative to the reference nucleotide sequence. For example, a nucleotide sequence that is 50%,
20 75%, 85%, 90%, 95% or 99% identical to a reference nucleotide sequence that is 100 bases long can have 50, 75, 85, 90, 95 or 99 bases that are completely identical to a 50, 75, 85, 90, 95 or 99 nucleotide sequence of the reference nucleotide sequence. The nucleotide sequence can also be a 100 base long nucleotide sequence that is 50%, 75%, 85%, 90%, 95% or 99% identical to the reference
25 nucleotide sequence over its entire length. Of course, there are other nucleotide sequences that will also meet the same criteria.

A nucleic acid sequence that is "substantially identical" to a VKOR nucleotide sequence is at least 80%, 85% 90%, 95% or 99% identical to the nucleotide sequence of SEQ ID NO:8 or 9. For purposes of comparison of nucleic acids, the
30 length of the reference nucleic acid sequence will generally be at least 40 nucleotides, e.g., at least 60 nucleotides or more nucleotides. Sequence identity can be measured using sequence analysis software (e.g., Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wis. 53705).

As is known in the art, a number of different programs can be used to identify whether a nucleic acid or amino acid has sequence identity or similarity to a known sequence. Sequence identity or similarity may be determined using standard techniques known in the art, including, but not limited to, the local sequence identity algorithm of Smith & Waterman, *Adv. Appl. Math.* 2, 482 (1981), by the sequence identity alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48,443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Natl. Acad. Sci. USA* 85,2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Drive, Madison, WI), the Best Fit sequence program described by Devereux *et al.*, *Nucl. Acid Res.* 12, 387-395 (1984), preferably using the default settings, or by inspection.

An example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments. It can also plot a tree showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng & Doolittle, *J. Mol. Evol.* 35, 351-360 (1987); the method is similar to that described by Higgins & Sharp, *CABIOS* 5, 151-153 (1989).

Another example of a useful algorithm is the BLAST algorithm, described in Altschul *et al.*, *J. Mol. Biol.* 215, 403-410, (1990) and Karlin *et al.*, *Proc. Natl. Acad. Sci. USA* 90, 5873-5787 (1993). A particularly useful BLAST program is the WU-BLAST-2 program that was obtained from Altschul *et al.*, *Methods in Enzymology*, 266, 460-480 (1996). WU-BLAST-2 uses several search parameters, which are preferably set to the default values. The parameters are dynamic values and are established by the program itself depending upon the composition of the particular sequence and composition of the particular database against which the sequence of interest is being searched; however, the values may be adjusted to increase sensitivity. An additional useful algorithm is gapped BLAST as reported by Altschul *et al.* *Nucleic Acids Res.* 25, 3389-3402.

The CLUSTAL program can also be used to determine sequence similarity. This algorithm is described by Higgins *et al.* (1988) *Gene* 73:237; Higgins *et al.* (1989) *CABIOS* 5:151-153; Corpet *et al.* (1988) *Nucleic Acids Res.* 16: 10881-90; Huang *et al.* (1992) *CABIOS* 8: 155-65; and Pearson *et al.* (1994) *Meth. Mol. Biol.* 24: 307-331.

In addition, for sequences that contain either more or fewer nucleotides than the nucleic acids disclosed herein, it is understood that in one embodiment, the percentage of sequence identity will be determined based on the number of identical nucleotides in relation to the total number of nucleotide bases. Thus, for example, sequence identity of sequences shorter than a sequence specifically disclosed herein will be determined using the number of nucleotide bases in the shorter sequence, in one embodiment. In percent identity calculations, relative weight is not assigned to various manifestations of sequence variation, such as, insertions, deletions, substitutions, etc.

10 The VKOR polypeptides of the invention include, but are not limited to, recombinant polypeptides, synthetic peptides and natural polypeptides. The invention also encompasses nucleic acid sequences that encode forms of VKOR polypeptides in which naturally occurring amino acid sequences are altered or deleted. Preferred nucleic acids encode polypeptides that are soluble under normal physiological conditions. Also within the invention are nucleic acids encoding fusion proteins in which all or a portion of VKOR is fused to an unrelated polypeptide (e.g., a marker polypeptide or a fusion partner) to create a fusion protein. For example, the polypeptide can be fused to a hexa-histidine tag to facilitate purification of bacterially expressed polypeptides, or to a hemagglutinin tag to facilitate purification of polypeptides expressed in eukaryotic cells, or to an HPC4 tag to facilitate purification of polypeptides by affinity chromatography or immunoprecipitation. The invention also includes isolated polypeptides (and the nucleic acids that encode these polypeptides) that include a first portion and a second portion; the first portion includes, e.g., all or a portion of a VKOR polypeptide, and the second portion includes, e.g., a detectable marker.

The fusion partner can be, for example, a polypeptide that facilitates secretion, e.g., a secretory sequence. Such a fused polypeptide is typically referred to as a preprotein. The secretory sequence can be cleaved by the cell to form the mature protein. Also within the invention are nucleic acids that encode VKOR fused to a polypeptide sequence to produce an inactive preprotein. Preproteins can be converted into the active form of the protein by removal of the inactivating sequence.

The invention also includes nucleic acids that hybridize, e.g., under stringent hybridization conditions (as defined herein) to all or a portion of the nucleotide sequence of SEQ ID NOS: 1-6, 8 or 9 or their complements. In particular

embodiments, the hybridizing portion of the hybridizing nucleic acid is typically at least 15 (e.g., 20, 30, or 50) nucleotides in length. The hybridizing portion of the hybridizing nucleic acid is at least 80%, e.g., at least 95%, at least 98% or 100%, identical to the sequence of a portion or all of a nucleic acid encoding a VKOR polypeptide. Hybridizing nucleic acids of the type described herein can be used, for example, as a cloning probe, a primer (e.g., a PCR primer), or a diagnostic probe. Also included within the invention are small inhibitory RNAs (siRNAs) and/or antisense RNAs that inhibit the function of VKOR, as determined, for example, in an activity assay, as described herein and as is known in the art.

10 In another embodiment, the invention features cells, e.g., transformed cells, that contain a nucleic acid of this invention. A "transformed cell" is a cell into which (or into an ancestor of which) has been introduced, by means of recombinant nucleic acid techniques, a nucleic acid encoding all or a part of a VKOR polypeptide, and/or an antisense nucleic acid or siRNA. Both prokaryotic and eukaryotic cells are included, e.g., bacteria, yeast, insect, mouse, rat, human, plant and the like.

The invention also features nucleic acid constructs (e.g., vectors and plasmids) that include a nucleic acid of the invention that is operably linked to a transcription and/or translation control elements to enable expression, e.g., expression vectors. By "operably linked" is meant that a selected nucleic acid, e.g., a DNA molecule encoding a VKOR polypeptide, is positioned adjacent to one or more regulatory elements, e.g., a promoter, which directs transcription and/or translation of the sequence such that the regulatory elements can control transcription and/or translation of the selected nucleic acid.

The present invention further provides fragments or oligonucleotides of the nucleic acids of this invention, which can be used as primers or probes. Thus, in some embodiments, a fragment or oligonucleotide of this invention is a nucleotide sequence that is at least 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 1000, 1500, 2000, 2500 or 3000 contiguous nucleotides of the nucleotide sequence set forth in SEQ ID NO:8 or SEQ ID NO:9. Examples of oligonucleotides of this invention are provided in the Sequence Listing included herewith. Such fragments or oligonucleotides can be detectably labeled or modified, for example, to include and/or incorporate a restriction enzyme cleavage site when employed as a primer in an amplification (e.g., PCR) assay.

The invention also features purified or isolated VKOR polypeptides, such as, for example, a polypeptide comprising, consisting essentially of and/or consisting of the amino acid sequence of SEQ ID NO:10 or a biologically active fragment or peptide thereof. Such fragments or peptides are typically at least about ten amino acids of the amino acid sequence of SEQ ID NO:10 (e.g., 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 75, 85, 95, 100, 125, or 150 amino acids of the amino acid sequence of SEQ ID NO:10) and can be peptides or fragment of contiguous amino acids of the amino acid sequence of the VKOR protein (e.g., as set forth in SEQ ID NO:10). The biological activity of a fragment or peptide of this invention can be determined according to the methods provided herein and as are known in the art for identifying VKOR activity. The fragments and peptides of the VKOR protein of this invention can also be active as antigens for the production of antibodies. The identification of epitopes on a fragment or peptide of this invention is carried out by well known protocols and would be within the ordinary skill of one in the art.

As used herein, both "protein" and "polypeptide" mean any chain of amino acids, regardless of length or post-translational modification (e.g., glycosylation, phosphorylation or N-myristylation). Thus, the term "VKOR polypeptide" includes full-length, naturally occurring VKOR proteins, respectively, as well as recombinantly or synthetically produced polypeptides that correspond to a full-length, naturally occurring VKOR protein, or to a portion of a naturally occurring or synthetic VKOR polypeptide.

A "purified" or "isolated" compound or polypeptide is a composition that is at least 60% by weight the compound of interest, e.g., a VKOR polypeptide or antibody that is separated or substantially free from at least some of the other components of the naturally occurring organism or virus, for example, the cell or viral structural components or other polypeptides or nucleic acids commonly found associated with the polypeptide. As used herein, the "isolated" polypeptide is at least about 25%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99% or more pure (w/w). Preferably the preparation is at least 75% (e.g., at least 90% or 99%) by weight the compound of interest. Purity can be measured by any appropriate standard method, e.g., column chromatography, polyacrylamide gel electrophoresis, or HPLC analysis.

Preferred VKOR polypeptides include a sequence substantially identical to all or a portion of a naturally occurring VKOR polypeptide. Polypeptides "substantially identical" to the VKOR polypeptide sequences described herein have an amino acid

sequence that is at least 80% or 85% (e.g., 90%, 95% or 99%) identical to the amino acid sequence of the VKOR polypeptides of SEQ ID NO: 10. For purposes of comparison, the length of the reference VKOR polypeptide sequence will generally be at least 16 amino acids, e.g., at least 20, 25, 30, 35, 40, 45, 50, 75, or 100 amino acids.

In the case of polypeptide sequences that are less than 100% identical to a reference sequence, the non-identical positions are preferably, but not necessarily, conservative substitutions for the reference sequence. Conservative substitutions typically include, but are not limited to, substitutions within the following groups:

10 glycine and alanine; valine, isoleucine, and leucine; aspartic acid and glutamic acid; asparagine and glutamine; serine and threonine; lysine and arginine; and phenylalanine and tyrosine.

Where a particular polypeptide is said to have a specific percent identity to a reference polypeptide of a defined length, the percent identity is relative to the reference polypeptide. Thus, for example, a polypeptide that is 50%, 75%, 85%, 90%, 95% or 99% identical to a reference polypeptide that is 100 amino acids long can be a 50, 75, 85, 90, 95 or 99 amino acid polypeptide that is completely identical to a 50, 75, 85, 90, 95 or 99 amino acid long portion of the reference polypeptide. It can also be a 100 amino acid long polypeptide that is 50%, 75%, 85%, 90%, 95% or 99% identical to the reference polypeptide over its entire length. Of course, other polypeptides also will meet the same criteria.

The invention also features purified or isolated antibodies that specifically bind to a VKOR polypeptide of this invention or to a fragment thereof. By "specifically binds" is meant that an antibody recognizes and binds a particular antigen, e.g., a VKOR polypeptide, or an epitope on a fragment or peptide of a VKOR polypeptide, but does not substantially recognize and bind other molecules in a sample. In one embodiment the antibody is a monoclonal antibody and in other embodiments, the antibody is a polyclonal antibody. The production of both monoclonal and polyclonal antibodies, including chimeric antibodies, humanized antibodies, single chain antibodies, bi-specific antibodies, antibody fragments, etc., is well known in the art.

In another aspect, the invention features a method for detecting a VKOR polypeptide in a sample. This method comprises contacting the sample with an antibody that specifically binds a VKOR polypeptide or a fragment thereof under conditions that allow the formation of a complex between an antibody and VKOR;

and detecting the formation of a complex, if any, as detection of a VKOR polypeptide or fragment thereof in the sample. Such immunoassays are well known in the art and include immunoprecipitation assays, immunoblotting assays, immunolabeling assays, ELISA, etc.

- 5 The present invention further provides a method of detecting a nucleic acid encoding a VKOR polypeptide in a sample, comprising contacting the sample with a nucleic acid of this invention that encodes VKOR or a fragment thereof, or a complement of a nucleic acid that encodes VKOR or a fragment thereof, under conditions whereby a hybridization complex can form, and detecting formation of a
- 10 hybridization complex; thereby detecting a nucleic acid encoding a VKOR polypeptide in a sample. Such hybridization assays are well known in the art and include probe detection assays and nucleic acid amplification assays.

- Also encompassed by the invention is a method of obtaining a gene related to (i.e., a functional homologue of) the VKOR gene. Such a method entails obtaining or
- 15 producing a detectably-labeled probe comprising an isolated nucleic acid which encodes all or a portion of VKOR, or a homolog thereof; screening a nucleic acid fragment library with the labeled probe under conditions that allow hybridization of the probe to nucleic acid fragments in the library, thereby forming nucleic acid duplexes; isolating labeled duplexes, if any; and preparing a full-length gene
- 20 sequence from the nucleic acid fragments in any labeled duplex to obtain a gene related to the VKOR gene.

- A further aspect of the present invention is a method of making a vitamin K dependent protein which comprises culturing a host cell that expresses a nucleic acid encoding a vitamin K dependent protein in the presence of vitamin K and
- 25 produces a vitamin K dependent protein, and then harvesting the vitamin K dependent protein from the culture, the host cell containing and expressing a heterologous nucleic acid encoding vitamin K dependent carboxylase, and the host cell further containing and expressing a heterologous nucleic acid encoding vitamin
- 30 K epoxide reductase (VKOR) and producing VKOR as described herein. The expression of the VKOR-encoding nucleic acid and the production of the VKOR causes the cell to produce greater levels of the vitamin K dependent protein than would be produced in the absence of the VKOR.

 Thus, in some embodiments, the present invention also provides a method of producing a vitamin K dependent protein, comprising:

a) introducing into a cell a nucleic acid that encodes a vitamin K dependent protein under conditions whereby the nucleic acid is expressed and the vitamin K dependent protein is produced in the presence of vitamin K, wherein the cell comprises a heterologous nucleic acid encoding vitamin K dependent carboxylase and further comprises a heterologous nucleic acid encoding vitamin K epoxide reductase; and

b) optionally collecting the vitamin K dependent protein from the cell. The vitamin K dependent protein that can be produced can be any vitamin K dependent protein now known or later identified as such, including but not limited to Factor VII, Factor IX, Factor X, Protein C, Protein S and prothrombin, in any combination. Any host cell that can be transformed with the nucleic acids described can be used as described herein, although in some embodiments non-human or even non-mammalian host cells can be used. Nucleic acids encoding vitamin K dependent carboxylase and nucleic acids encoding vitamin K dependent proteins as described herein are well known in the art and their introduction into cells for expression would be carried out according to routine protocols.

Certain embodiments of this invention are based on the inventors' discovery that a subject's therapeutic dose of warfarin for anticoagulation therapy can be correlated with the presence of one or more single nucleotide polymorphisms in the VKOR gene of the subject. Thus, the present invention also provides a method of identifying a human subject having increased or decreased sensitivity to warfarin, comprising detecting in the subject the presence of a single nucleotide polymorphism (SNP) in the VKOR gene, wherein the single nucleotide polymorphism is correlated with increased or decreased sensitivity to warfarin, thereby identifying the subject as having increased or decreased sensitivity to warfarin.

An example of a SNP correlated with an increased sensitivity to warfarin is a G→C alteration at nucleotide 2581 (SEQ ID NO:12) (in intron 2 of the VKOR gene; GenBank accession no. refSNP ID: rs8050894, incorporated by reference herein) of the nucleotide sequence of SEQ ID NO:11, which is a reference sequence encompassing the genomic sequence of SEQ ID NO:8 and approximately 1000 nucleotides preceding and following this sequence. This sequence can be located as having the genome position "human chromosome 16p11.2" or in the physical map in the NCBI database as human chromosome 16: 31009700-31013800.

Examples of SNPs correlated with a decreased sensitivity to warfarin are a T→C alteration at nucleotide 3294 (SEQ ID NO:13) (in intron 2 of the VKOR gene; GenBank accession no. refSNP ID: rs2359612, incorporated by reference herein) of the nucleotide sequence of SEQ ID NO:11 and a G→A alteration at nucleotide 4769 (SEQ ID NO:14) (in the 3' UTR of the VKOR gene; GenBank accession no. refSNP ID: rs7294, incorporated by reference herein) of the nucleotide sequence of SEQ ID NO:11.

As used herein, a subject having an "increased sensitivity to warfarin" is a subject for whom a suitable therapeutic or maintenance dose of warfarin is lower than the therapeutic or maintenance dose of warfarin that would be suitable for a normal subject, i.e., a subject who did not carry a SNP in the VKOR gene that imparts a phenotype of increased sensitivity to warfarin. Conversely, as used herein, a subject having a "decreased sensitivity to warfarin" is a subject for whom a suitable therapeutic or maintenance dose of warfarin is higher than the therapeutic or maintenance dose of warfarin that would be suitable for a normal subject, i.e., a subject who did not carry a SNP in the VKOR gene that imparts a phenotype of decreased sensitivity to warfarin. An example of a typical therapeutic dose of warfarin for a normal subject is 35 mg per week, although this amount can vary (e.g., a dose range of 3.5 to 420 mg per week is described in Aithal et al. (1999) *Lancet* 353:717-719). A typical therapeutic dose of warfarin can be determined for a given study group according to the methods described herein, which can be used to identify subjects with therapeutic warfarin doses above or below this dose, thereby identifying subjects having decreased or increased sensitivity to warfarin.

Further provided herein is a method of identifying a human subject having increased or decreased sensitivity to warfarin, comprising: a) correlating the presence of a single nucleotide polymorphism in the VKOR gene with increased or decreased sensitivity to warfarin; and b) detecting the single nucleotide polymorphism of step (a) in the subject, thereby identifying a subject having increased or decreased sensitivity to warfarin.

In addition, the present invention provides a method of identifying a single nucleotide polymorphism in the VKOR gene correlated with increased or decreased sensitivity to warfarin, comprising: a) identifying a subject having increased or decreased sensitivity to warfarin; b) detecting in the subject the presence of a single nucleotide polymorphism in the VKOR gene; and c) correlating the presence of the

single nucleotide polymorphism of step (b) with the increased or decreased sensitivity to warfarin in the subject, thereby identifying a single nucleotide polymorphism in the VKOR gene correlated with increased or decreased sensitivity to warfarin.

- 5 Also provided herein is a method of correlating a single nucleotide polymorphism in the VKOR gene of a subject with increased or decreased sensitivity to warfarin, comprising: a) identifying a subject having increased or decreased sensitivity to warfarin; b) determining the nucleotide sequence of the VKOR gene of the subject of (a); c) comparing the nucleotide sequence of step (b) with the wild type
10 nucleotide sequence of the VKOR gene; d) detecting a single nucleotide polymorphism in the nucleotide sequence of (b); and e) correlating the single nucleotide polymorphism of (d) with increased or decreased sensitivity to warfarin in the subject of (a).

- A subject is identified as having an increased or decreased sensitivity to
15 warfarin by establishing a therapeutic or maintenance dose of warfarin for anticoagulation therapy according to well known protocols and comparing the therapeutic or maintenance dose for that subject with the therapeutic or maintenance dose of warfarin for anticoagulation therapy of a population of normal subjects (e.g., subjects lacking any SNPs in the VKOR gene correlated with increased or
20 decreased sensitivity to warfarin) from which an average or mean therapeutic or maintenance dose of warfarin is calculated. A subject having a therapeutic or maintenance dose of warfarin that is below the average therapeutic or maintenance dose of warfarin (e.g., the dose of warfarin that is therapeutic or provides a maintenance level for a subject that has a wild type VKOR gene, i.e., lacking any
25 single nucleotide polymorphisms associated with warfarin sensitivity) is a subject identified as having an increased sensitivity to warfarin. A subject having a therapeutic or maintenance dose of warfarin that is above the average therapeutic or maintenance of warfarin is a subject identified as having a decreased sensitivity to warfarin. An average therapeutic or maintenance dose of warfarin for a subject with
30 a wild type VKOR gene would be readily determined by one skilled in the art.

 The nucleotide sequence of the VKOR gene of a subject is determined according to methods standard in the art, and as described in the Examples provided herein. For example, genomic DNA is extracted from cells of a subject and the VKOR gene is located and sequenced according to known protocols. Single

nucleotide polymorphisms in the VKOR gene are identified by a comparison of a subject's sequence with the wild type sequence as known in the art (e.g., the reference sequence as shown herein as SEQ ID NO:11).

5 A SNP in the VKOR gene is correlated with an increased or decreased sensitivity to warfarin by identifying the presence of a SNP or multiple SNPs in the VKOR gene of a subject also identified as having increased or decreased sensitivity to warfarin, i.e., having a maintenance or therapeutic dose of warfarin that is above or below the average dose and performing a statistical analysis of the association of the SNP or SNPs with the increased or decreased sensitivity to warfarin, according to well
10 known methods of statistical analysis. An analysis that identifies a statistical association (e.g., a significant association) between the SNP(s) (genotype) and increased or decreased warfarin sensitivity (phenotype) establishes a correlation between the presence of the SNP(s) in a subject and an increased or decreased sensitivity to warfarin in that subject.

15 It is contemplated that a combination of factors, including the presence of one or more SNPs in the VKOR gene of a subject, can be correlated with an increased or decreased sensitivity to warfarin in that subject. Such factors can include, but are not limited to cytochrome p450 2C9 polymorphisms, race, age, gender, smoking history and hepatic disease.

20 Thus, in a further embodiment, the present invention provides a method of identifying a human subject having increased or decreased sensitivity to warfarin, comprising identifying in the subject the presence of a combination of factors correlated with an increased or decreased sensitivity to warfarin selected from the group consisting of one or more single nucleotide polymorphisms of the VKOR gene,
25 one or more cytochrome p450 2C9 polymorphisms, race, age, gender, smoking history, hepatic disease and any combination of two or more of these factors, wherein the combination of factors is correlated with increased or decreased sensitivity to warfarin, thereby identifying the subject having increased or decreased sensitivity to warfarin.

30 Further provided herein is a method of identifying a human subject having increased or decreased sensitivity to warfarin, comprising: a) correlating the presence of a combination of factors with an increased or decreased sensitivity to warfarin, wherein the factors are selected from the group consisting of one or more single nucleotide polymorphisms of the VKOR gene, one or more cytochrome p450

2C9 polymorphisms, race, age, gender, smoking history, hepatic disease and any combination of two or more of these factors; and b) detecting the combination of factors of step (a) in the subject, thereby identifying a subject having increased or decreased sensitivity to warfarin.

5 In addition, the present invention provides a method of identifying a combination of factors correlated with an increased or decreased sensitivity to warfarin, wherein the factors are selected from the group consisting of one or more single nucleotide polymorphisms of the VKOR gene, one or more cytochrome p450 2C9 polymorphisms, race, age, gender, smoking history, hepatic disease and any
10 combination of two or more of these factors; comprising: a) identifying a subject having increased or decreased sensitivity to warfarin; b) detecting in the subject the presence of a combination of the factors; and c) correlating the presence of the combination of factors of step (b) with the increased or decreased sensitivity to warfarin in the subject, thereby identifying a combination of factors correlated with
15 increased or decreased sensitivity to warfarin.

Also provided herein is a method of correlating a combination of factors, wherein the factors are selected from the group consisting of one or more single nucleotide polymorphisms of the VKOR gene, one or more cytochrome p450 2C9 polymorphisms, race, age, gender, smoking history, hepatic disease and any
20 combination of two or more of these factors, with increased or decreased sensitivity to warfarin, comprising: a) identifying a subject having increased or decreased sensitivity to warfarin; b) identifying the presence of a combination of the factors in the subject; and c) correlating the combination of the factors of (b) with increased or decreased sensitivity to warfarin in the subject of (a).

25 A combination of factors as described herein is correlated with an increased or decreased sensitivity to warfarin by identifying the presence of the combination of factors in a subject also identified as having increased or decreased sensitivity to warfarin and performing a statistical analysis of the association of the combination of factors with the increased or decreased sensitivity to warfarin, according to well
30 known methods of statistical analysis. An analysis that identifies a statistical association (e.g., a significant association) between the combination of factors and the warfarin sensitivity phenotype (increased or decreased) establishes a correlation between the presence of the combination of factors in a subject and an increased or decreased sensitivity to warfarin in that subject.

Further provided herein are nucleic acids encoding VKOR and comprising one or more SNPs as described herein. Thus, the present invention further provides nucleic acids comprising, consisting essentially of and/or consisting of the nucleotide sequence as set forth in SEQ ID NOs:12, 13, 14, 15 and 16. The nucleic acids can be present in a vector and the vector can be present in a cell. Further included are proteins encoded by a nucleic acid comprising a nucleotide sequence as set forth in SEQ ID NOs:12, 13, 14, 15 and 16, as well as antibodies that specifically bind a protein encoded by a nucleic acid comprising a nucleotide sequence as set forth in SEQ ID NOs:12, 13, 14, 15 and 16. The present invention is more particularly described in the following examples that are intended as illustrative only since numerous modifications and variations therein will be apparent to those skilled in the art.

EXAMPLES

15

EXAMPLE I CORRELATION BETWEEN SNPS IN VKOR GENE AND INCREASED OR DECREASED SENSITIVITY TO WARFARIN

The most prevalent isoform of the VKOR gene is about 4 kb long, has three exons and encodes an enzyme of 163 amino acids with a mass of 18.4 kDa. In the present study, three mutations vk2581(G>C), vk3294(T>C) and vk4769(G>A), identified as SNPs (heterozygosity ratios of 46.9%, 46.8% and 46.3%, respectively) were examined for a correlation between their presence in a subject and the maintenance dose of warfarin required to achieve a therapeutically effective response.

25 1. Selection of subjects

Subjects were obtained from the UNC Coagulation Clinic in the Ambulatory Care Center. Informed consent was obtained by a trained genetic counselor. Subjects not fluent in English were excluded because of the lack of translators and the requirement for consent. To qualify for the study, subjects had warfarin for at least six months, were older than 18 and were followed by the UNC Coagulation clinic at the Ambulatory Care Clinic.

30

2. Extraction of genomic DNA from whole blood

Genomic DNAs were extracted from the whole blood of subjects using QIAamp DNA Blood Mini Kit (QIAGEN cat#51104). The DNA concentration was adjusted to 10 ng/ μ L.

3. Sequencing of the genomic DNA samples

5 Approximately 10 ng of DNA was used for polymerase chain reaction (PCR) assays. The primers used to amplify the VKOR gene were: Exon 1-5' CCAATCGCCGAGTCAGAGG (SEQ ID NO:29) and Exon 1-3' CCCAGTCCCCAGCACTGTCT (SEQ ID NO:30) for the 5'-UTR and Exon 1 region; Exon 2-5' AGGGGAGGATAGGGTCACTG (SEQ ID NO:31) and Exon 2-3' 10 --- CCTGTTAGTTACCTCCCCACA (SEQ ID NO:32) for the Exon 2 region; and Exon 3-5' ATACGTGCGTAAGCCACCAC (SEQ ID NO:33) and Exon 3-3' ACCCAGATATGCCCCCTTAG (SEQ ID NO:34) for the Exon3 and 3'-UTR region. Automated high throughput capillary electrophoresis DNA sequencing was used for detecting SNPs in the VKOR gene.

15 4. Detection of known SNPs using real-time PCR

 The assay reagents for SNP genotyping were from the Assay-by-Design™ service (Applied Biosystems, cat#4332072). The primers and probes (FAM™ and VIC™ dye-labeled) were designed using Primer Express software and were synthesized in an Applied Biosystems synthesizer. The primer pairs for each SNP 20 are located at the upstream/downstream position of the SNP site and can generate less than 100 bp length of a DNA fragment in the PCR reaction. The FAM™ and VIC™ dye-labeled probes were designed to cover the SNP sites with a length of 15-16 nt. The primer and probe sequences for each VKOR SNP are shown in Table 2.

 The 2X TaqMan™ Universal PCR Master Mix, No AmpErase UNG (Applied 25 Biosystems, cat#4324018) was used in the PCR reactions. Forty cycles of real-time PCR were performed in an Opticon II (MJ Research) machine. There was a 10 minute 95°C preheat followed by 92°C for 15 sec, 60°C for 1 min. and then a plate reading. The results were read according to the signal value of FAM and VIC dye.

5. Statistical analysis

30 The difference of average dose between different genotypes was compared by analysis of variance (ANOVA) using SAS version 8.0 (SAS, Inc., Cary, NC). A two-sided p value less than 0.05 was considered significant. Examination of the distribution and residuals for the average dose of treatment among the SNP groups

indicated that a log transformation was necessary to satisfy the assumption of homogeneity of variance.

6. Correlation of SNPs with warfarin dosage

By direct genomic DNA sequencing and SNP real-time PCR detection, five
5 SNPs were identified in the VKOR gene: one in the 5'-UTR, two in intron II, one in the coding region and one in the 3'-UTR (Table 1).

Among these SNPs, the vk563 and vk4501 SNPs allele were carried by only one of the 58 subjects of the study (a triple heterozygous, also carrying the 3'-UTR SNP allele), while the other SNPs were identified in 17-25 heterozygous patients.

10 Each marker was first analyzed independently. Figure 1A shows that the average warfarin dose for patients with the vk2581 wild type allele was 50.19 ± 3.20 mg per week ($n=26$), while those heterozygous and homozygous for this polymorphism were 35.19 ± 3.73 ($n=17$) and 31.14 ± 6.2 mg per week ($n=15$), respectively. Figure 1B shows that the average warfarin dose for patients with the
15 wild-type vk3294 allele was 25.29 ± 3.05 mg per week ($n=11$), while patients bearing the heterozygous and homozygous alleles were 41.68 ± 4.92 ($n=25$) and 47.73 ± 2.75 mg per week ($n=22$), respectively. Figure 1C shows the average warfarin dose for patients with vk4769 SNP wild type was 35.35 ± 4.01 mg per week ($n=27$), while patients with the heterozygous and homozygous alleles required 44.48 ± 4.80 ($n=19$)
20 and 47.56 ± 3.86 mg per week ($n=12$), respectively. It was also observed that P450 2C9 *3 has a significant effect on warfarin dose (Figure 1D), as previously reported (Joffe et al. (2004) "Warfarin dosing and cytochrome P450 2C9 polymorphisms" *Thromb Haemost* 91:1123-1128). The average warfarin dose for patients with P450 2C9 *1 (wild type) was 43.82 ± 2.75 mg per week ($n=50$), while patients heterozygous
25 for this allele required 22.4 ± 4.34 mg per week ($n=8$).

7. Statistical analysis

The association of the $\text{Log}_e(\text{warfarin average dosage})$ (LnDose) with the SNPs in the VKOR gene was examined by analysis of variance (ANOVA). SAS was used first to do a repeated procedure in which a series of factors (race, gender,
30 smoking history, hepatic diseases, SNPs at cytochrome P450 2Y9 gene, etc.) were examined to identify factors, excluding VKOR SNPs, which might affect dosage. P450 2C9 *3 was significantly associated with the average dose of warfarin; thus, it was included as a covariant for further analysis. The analysis indicated that the three VKOR SNPs were still significantly associated with weekly warfarin dose

(vk2581, $P < 0.0001$; vk3294, $P < 0.0001$; and vk4769, $P = 0.0044$), when the covariance is included.

To specifically test if the three SNPs of VKOR were independently associated with warfarin dosage, the analysis was repeated in which two SNPs in the VKOR gene were included as covariates for the other SNP. The three VKOR SNPs are located within 2 kb distance of one another and are expected to be closely linked. It was clear from inspection that, at least for Caucasians, one haplotype (where A=vk2581 guanine and a=vk2581 cytosine; B=vk3294 thymine and b=vk3294 cytosine; C=vk4769 guanine and c=vk4769 adenine) was AAbbcc and another aaBBCC. The distribution of individual SNPs in patients was found to be significantly correlated with the others ($R=0.63-0.87$, $p<0.001$). Indeed, subjects with the haplotype AAbbcc ($n=7$) required a significantly higher dosage of warfarin (warfarin dosage= 48.98 ± 3.93) compared to those patients with haplotype aaBBCC (25.29 ± 3.05 ; $p<0.001$).

EXAMPLE 2 siRNA DESIGN AND SYNTHESIS

siRNAs were selected using an advanced version of a rational design algorithm (Reynolds et al. (2004) "Rational siRNA design for RNA interference" *Nature Biotechnology* 22:326-330). For each of the 13 genes, four siRNAs duplexes with the highest scores were selected and a BLAST search was conducted using the Human EST database. To minimize the potential for off-target silencing effects, only those sequence targets with more than three mismatches against un-related sequences were selected (Jackson et al. (2003) "Expression profiling reveals off-target gene regulation by RNAi" *Nat Biotechnol* 21:635-7). All duplexes were synthesized in Dharmacon (Lafayette, CO) as 21-mers with UU overhangs using a modified method of 2'-ACE chemistry (Scaringe (2000) "Advanced 5'-silyl-2'-orthoester approach to RNA oligonucleotide synthesis" *Methods Enzymol* 317:3-18) and the AS strand was chemically phosphorylated to ensure maximum activity (Martinez et al. (2002) "Single-stranded antisense siRNAs guide target RNA cleavage in RNAi" *Cell* 110:563-74).

EXAMPLE 3 siRNA transfection

Transfection was essentially as previously described (Harborth et al. (2001) "Identification of essential genes in cultured mammalian cells using small interfering RNAs" *J Cell Sci* 114:4557-65) with minor modifications.

5

EXAMPLE 4 VKOR activity assay

siRNA transfected A549 cells were trypsinized and washed twice with cold PBS. 1.5×10^7 cells were taken for each VKOR assay. 200 μ L buffer D (250 mM $\text{Na}_2\text{HPO}_4\text{-NaH}_2\text{PO}_4$, 500 mM KCl, 20% glycerol and 0.75% CHAPS, pH 7.4) was added to the cell pellet, followed by sonication of the cell lysate. For assays of solubilized microsomes, microsomes were prepared from 2×10^9 cells as described (Lin et al. (2002) "The putative vitamin K-dependent gamma-glutamyl carboxylase internal propeptide appears to be the propeptide binding site" *J Biol Chem* 277:28584-91); 10 to 50 μ L of solubilized microsomes were used for each assay. Vitamin K epoxide was added to the concentration indicated in the figure legends and DTT was added to 4 mM to initiate the reaction. The reaction mixture was incubated in yellow light at 30°C for 30 minutes and stopped by adding 500 μ L 0.05 M AgNO_3 : isopropanol (5:9). 500 μ L hexane was added and the mixture was vortexed vigorously for 1 minute to extract the vitamin K and KO. After 5 minutes centrifugation, the upper organic layer was transferred to a 5-mL brown vial and dried with N_2 . 150 μ L HPLC buffer, acetonitrile:isopropanol:water (100:7:2), was added to dissolve the vitamin K and KO and the sample was analyzed by HPLC on an A C-18 column (Vydac, cat#218TP54).

EXAMPLE 5 RT-qPCR (reverse transcriptase quantitative PCR)

1×10^6 cells were washed with PBS twice and total RNA was isolated with Trizol reagent according to the manufacturer's protocol (Invitrogen). 1 μ g of RNA was digested by RQ1 DNaseI (Promega) and heat-inactivated. First strand cDNA was made with M-MLV reverse transcriptase (Invitrogen). cDNAs were mixed with DyNAmo SYBR Green qPCR pre-mix (Finnzymes) and real-time PCR was performed with an Opticon II PCR thermal cycler (MJ Research). The following primers were used:

13124769-5' (F): (TCCAACAGCATATTCGGTTGC, SEQ ID NO: 1);

13124769-3 (R)': (TTCTTGGACCTTCCGGAACT, SEQ ID NO: 2);

GAPDH-F: (GAAGGTGAAGGTCGGAGTC, SEQ ID NO: 3);
GAPDH-R: (GAAGATGGTGATGGGATTTC, SEQ ID NO: 4);
Lamin-RT-F: (CTAGGTGAGGCCAAGAAGCAA, SEQ ID NO: 5) and
Lamin-RT-R: (CTGTTCCTCTCAGCAGACTGC, SEQ ID NO: 6).

5

EXAMPLE 6 Over-expression of VKOR in Sf9 insect cell line

The cDNA for the mGC11276 coding region was cloned into pVL1392 (Pharmingen), with the HPC4 tag (EDQVDPRLIDGK, SEQ ID NO: 7) at its amino terminus and expressed in Sf9 cells as described (Li et al. (2000) "Identification of a
10 *Drosophila* vitamin K-dependent gamma-glutamyl carboxylase" *J Biol Chem* 275:18291-6).

EXAMPLE 7 Gene selection

The search for the VKOR gene was focused on human chromosome sixteen
15 between markers D16S3131 and D16S419. This region corresponds to chromosome 16 at 50cM-65cM on the genetic map and 26-46.3Mb on the physical map. 190 predicted coding sequences in this region were analyzed by a BLASTX search of the NCBI non-redundant protein database. Those human genes and orthologs from related species with known function were eliminated. Because VKOR appears to be
20 a transmembrane protein (Carlisle & Suttie (1980) "Vitamin K dependent carboxylase: subcellular location of the carboxylase and enzymes involved in vitamin K metabolism in rat liver" *Biochemistry* 19:1161-7), the remaining genes were translated according to the cDNA sequences in the NCBI database and analyzed with the programs TMHMM and TMAP (Biology WorkBench, San Diego
25 Supercomputer System) to predict those with transmembrane domains. Thirteen genes predicted to code for integral membrane proteins were chosen for further analysis.

EXAMPLE 8 Cell line screening for VKOR activity

30 The strategy was to identify a cell line expressing relatively high amounts of VKOR activity and use siRNA to systematically knock down all thirteen candidate genes. siRNA, double stranded RNA of 21-23 nucleotides, has been shown to cause specific RNA degradation in cell culture (Hara et al. (2002) "Raptor, a binding partner of target of rapamycin (TOR), mediates TOR action" *Cell* 110:177-89; Krichevsky &

Kosik (2002) "RNAi functions in cultured mammalian neurons" *Proc Natl Acad Sci USA* 99:11926-9; Burns et al. (2003) "Silencing of the Novel p53 Target Gene Snk/Plk2 Leads to Mitotic Catastrophe in Paclitaxel (Taxol)-Exposed Cells" *Mol Cell Biol* 23:5556-71). However, application of siRNA for large scale screening in
5 mammalian cells has not previously been reported because of the difficulty in identifying a functional target for a specific mammalian cell mRNA (Holen et al. (2003) "Similar behaviour of single-strand and double-strand siRNAs suggests they act through a common RNAi pathway" *Nucleic Acids Res* 31:2401-7). The development of a rational selection algorithm (Reynolds et al.) for siRNA design
10 increases the probability that a specific siRNA can be developed; furthermore, the probability of success can be increased by pooling four rationally selected siRNAs. Using siRNA to search for previously unidentified genes has the advantage that, even if VKOR activity requires the product of more than one gene for activity, the screen should still be effective because the assay determines the loss of enzymatic
15 activity.

Fifteen cell lines were screened and a human lung carcinoma line, A549, was identified to exhibit sufficient warfarin-sensitive VKOR activity for facile measurement. A second human colorectal adenocarcinoma cell line, HT29, which expressed very little VKOR activity, was used as a reference.

20

EXAMPLE 9 siRNA inhibition of VKOR activity in A549 cells

Each of the thirteen pools of siRNA were transfected in triplicate into A549 cells and assayed for VKOR activity after 72 hours. One siRNA pool specific for gene gi:13124769 reduced VKOR activity by 64%-70% in eight separate assays
25 (Fig. 2).

One possible reason that VKOR activity was inhibited to only ~35% of its initial activity after 72 hours is that the half-life of mammalian proteins varies greatly (from minutes to days) (Zhang et al. (1996) "The major calpain isozymes are long-lived proteins. Design of an antisense strategy for calpain depletion in cultured cells" *J Biol Chem* 271:18825-30; Bohley (1996) "Surface hydrophobicity and intracellular degradation of proteins" *Biol Chem* 377:425-35; Dice & Goldberg (1975) "Relationship between in vivo degradative rates and isoelectric points of proteins" *Proc Natl Acad Sci USA* 72:3893-7), and mRNA translation is being inhibited, not
30 enzyme activity. Therefore, the cells were carried through eleven days and their

VKOR activity followed. Figure 3 shows that the level of mRNA for gi:13124769 mRNA decreased rapidly to about 20% of normal while VKOR activity decreased continuously during this time period. This reduction in activity is not a general effect of the siRNA or the result of cell death because the level of VKD carboxylase activity and lamin A/C mRNA remained constant. Furthermore, the level of gi:13124769 mRNA is four fold lower in HT-29 cells, which have low VKOR activity, than in A549 cells that exhibit high VKOR activity. These data indicate that gi:13124769 corresponds to the VKOR gene.

10 **EXAMPLE 10 Identification of gene encoding VKOR**

The gene, IMAGE 3455200 (gi:13124769, **SEQ ID NO: 8**), identified herein to encode VKOR, maps to human chromosome 16p11.2, mouse chromosome 7F3, and rat chromosome 1:180.8 Mb. There are 338 cDNA clones in the NCBI database representing seven different splicing patterns (NCBI AceView program). These are composed of all or part of two to four exons. Among these, the most prevalent isoform, mGC11276, has three exons and is expressed at high levels in lung and liver cells. This three exon transcript (**SEQ ID NO: 9**) encodes a predicted protein of 163 amino acids with a mass of 18.2 kDa (**SEQ ID NO: 10**). It is a putative N-myristylated endoplasmic reticulum protein with one to three transmembrane domains, depending upon the program used for prediction. It has seven cysteine residues, which is consistent with observations that the enzymatic activity is dependent upon thiol reagents (Thijssen et al. (1994) "Microsomal lipoamide reductase provides vitamin K epoxide reductase with reducing equivalents" *Biochem J* 297:277-80). Five of the seven cysteines are conserved among human, mice, rat, zebrafish, *Xenopus* and *Anopheles*.

To confirm that the VKOR gene had been identified, the most prevalent form of the enzyme (three exons) was expressed in *Spodoptera frugiperda*, Sf9 cells. Sf9 cells have no measurable VKOR activity but exhibit warfarin sensitive activity when transfected with mGC11276 cDNA (Figure 4). VKOR activity is observed from constructs with an epitope tag at either their amino or carboxyl terminus. This tag should assist in the purification of VKOR.

VKOR should exhibit warfarin sensitivity, therefore microsomes were made from Sf9 cells expressing VKOR and tested for warfarin sensitivity. The VKOR activity is warfarin-sensitive (Figure 5).

In summary, the present invention provides the first example of using siRNA in mammalian cells to identify an unknown gene. The identity of the VKOR gene was confirmed by its expression in insect cells. The VKOR gene encodes several isoforms. It will be important to characterize the activity and expression pattern of each isoform. Millions of people world-wide utilize warfarin to inhibit coagulation; therefore it is important to further characterize VKOR as it can lead to more accurate dosing or design of safer, more effective, anti-coagulants.

The foregoing is illustrative of the present invention, and is not to be construed as limiting thereof. The invention is defined by the following claims, with equivalents of the claims to be included therein.

All publications, patent applications, patents, patent publications and other references cited herein are incorporated by reference in their entireties for the teachings relevant to the sentence and/or paragraph in which the reference is presented.

Table 1. Five SNPs examined in VKOR gene

SNPs	position	AA change	Heterozygous ratio
vk563 G>A (SEQ ID NO:15)	5'- UTR	N/A	1/58
vk2581 G>C (SEQ ID NO:12)	Intro n2	N/A	17/58
vk3294 T>C (SEQ ID NO:13)	Intro n2	N/A	25/58
vk4501 C>T (SEQ ID NO:16)	Exo n3	Leu120Leu	1/58
vk4769 G>A (SEQ ID NO:14)	3'- UTR	N/A	19/58

Attorney Docket No. 5470.401WO

Table 2.

SNPs	VIC Probe Sequence	FAM Probe Sequence	Forward Primer	Reverse Primer
vk2581 G>C	TCATCACGGAGCGTC (SEQ ID NO:17)	TCATCACCGAGCGTC (SEQ ID NO:18)	GGTGATCCACACAGCTGACA (SEQ ID NO:19)	CCTGTTAGTTACCTCCCCACATC (SEQ ID NO:20)
vk3294 T>C	CCAGGACCATGGTGC (SEQ ID NO:21)	CCAGGACCGTGGTGC (SEQ ID NO:22)	GCTCCAGAGAAGGCATCACT (SEQ ID NO:23)	GCCAAGTCTGAACCATGTGTCA (SEQ ID NO:24)
vk4769 G>A	ATACCCGCACATGAC (SEQ ID NO:25)	CATACCCACACATGAC (SEQ ID NO:26)	GTCCCTAGAAGGCCCTAGATGT (SEQ ID NO:27)	GTGTGGCACATTTGGTCCATT (SEQ ID NO:28)

THAT WHICH IS CLAIMED IS:

1. A method of identifying a human subject having increased sensitivity to warfarin, comprising detecting in the subject the presence of a single nucleotide polymorphism in the VKOR gene, wherein the single nucleotide polymorphism is correlated with increased sensitivity to warfarin; thereby identifying the subject having increased sensitivity to warfarin.

2. The method of claim 1, wherein the single nucleotide polymorphism in the VKOR gene is a G→C alteration at nucleotide 2581 of the nucleotide sequence of SEQ ID NO:11.

3. A method of identifying a human subject having increased sensitivity to warfarin, comprising:

a) correlating the presence of a single nucleotide polymorphism in the VKOR gene with increased sensitivity to warfarin; and

b) detecting the single nucleotide polymorphism of step (a) in the subject, thereby identifying a subject having increased sensitivity to warfarin.

4. A method of identifying a single nucleotide polymorphism in the VKOR gene correlated with increased sensitivity to warfarin, comprising:

a) identifying a subject having increased sensitivity to warfarin;

b) detecting in the subject the presence of a single nucleotide polymorphism in the VKOR gene; and

c) correlating the presence of the single nucleotide polymorphism of step (b) with the increased sensitivity to warfarin in the subject, thereby identifying a single nucleotide polymorphism in the VKOR gene correlated with increased sensitivity to warfarin.

5. A method of correlating a single nucleotide polymorphism in the VKOR gene of a subject with increased sensitivity to warfarin, comprising:

a) identifying a subject having increased sensitivity to warfarin;

- b) determining the nucleotide sequence of the VKOR gene of the subject of (a);
- c) comparing the nucleotide sequence of step (b) with the wild type nucleotide sequence of the VKOR gene;
- d) detecting a single nucleotide polymorphism in the nucleotide sequence of (b); and
- e) correlating the single nucleotide polymorphism of (d) with increased sensitivity to warfarin in the subject of (a).

6. A method of identifying a human subject having decreased sensitivity to warfarin, comprising detecting in the subject the presence of a single nucleotide polymorphism in the VKOR gene, wherein the single nucleotide polymorphism is correlated with decreased sensitivity to warfarin, thereby identifying the subject having decreased sensitivity to warfarin.

7. The method of claim 6, wherein the single nucleotide polymorphism in the VKOR gene is a T→C alteration at nucleotide 3294 of the nucleotide sequence of SEQ ID NO:11.

8. The method of claim 6, wherein the single nucleotide polymorphism in the VKOR gene is a G→A alteration at nucleotide 4769 of the nucleotide sequence of SEQ ID NO:11.

9. A method of identifying a human subject having decreased sensitivity to warfarin, comprising:

- a) correlating the presence of a single nucleotide polymorphism in the VKOR gene with decreased sensitivity to warfarin; and
- b) detecting the single nucleotide polymorphism of step (a) in the subject, thereby identifying a subject having decreased sensitivity to warfarin.

10. A method of identifying a single nucleotide polymorphism in the VKOR gene correlated with decreased sensitivity to warfarin, comprising:

- a) identifying a subject having decreased sensitivity to warfarin;

b) detecting in the subject the presence of a single nucleotide polymorphism in the VKOR gene; and

c) correlating the presence of the single nucleotide polymorphism of step (b) with the decreased sensitivity to warfarin in the subject, thereby identifying a single nucleotide polymorphism in the VKOR gene correlated with decreased sensitivity to warfarin.

11. A method of correlating a single nucleotide polymorphism in the VKOR gene of a subject with decreased sensitivity to warfarin, comprising:

a) identifying a subject having decreased sensitivity to warfarin;

b) determining the nucleotide sequence of the VKOR gene of the subject of (a);

c) comparing the nucleotide sequence of step (b) with the wild type nucleotide sequence of the VKOR gene;

d) detecting a single nucleotide polymorphism in the nucleotide sequence of (b); and

e) correlating the single nucleotide polymorphism of (d) with decreased sensitivity to warfarin in the subject of (a).

12. In a method of making a vitamin K dependent protein which comprises

a) culturing a host cell which expresses a nucleic acid encoding a vitamin K dependent protein in the presence of vitamin K and produces a vitamin K dependent protein, and

b) harvesting said vitamin K dependent protein from the culture, said host cell containing and expressing a heterologous nucleic acid encoding vitamin K dependent carboxylase, the improvement comprising:

employing as said host cell a host cell further containing and expressing a heterologous nucleic acid encoding vitamin K epoxide reductase (VKOR).

13. The method of claim 12, wherein said vitamin K dependent protein is selected from the group consisting of Factor VII, Factor IX, Factor X, Protein C, Protein S, and prothrombin.

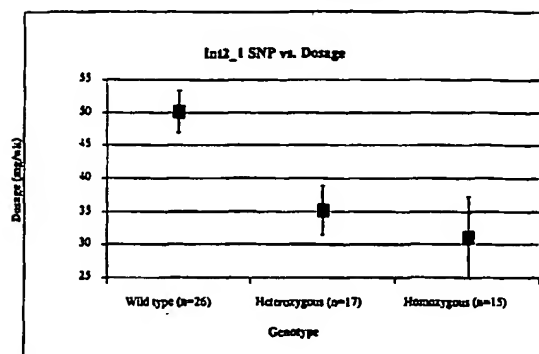
14. The method of claim 12, wherein said host cell is a plant cell.

15. The method of claim 12, wherein said host cell is an insect cell.

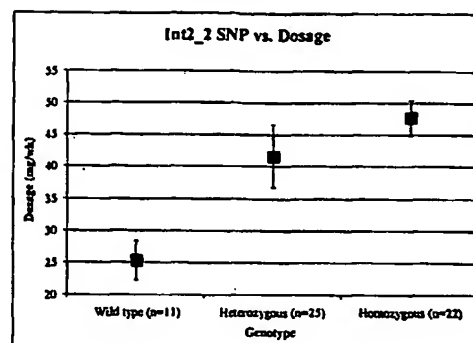
16. The method of claim 12, wherein said vitamin K-dependent carboxylase is bovine vitamin K dependent carboxylase.

1/3

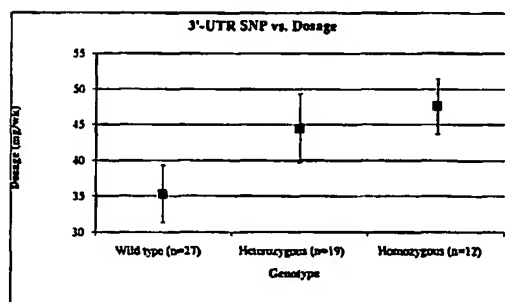
A. vk2581



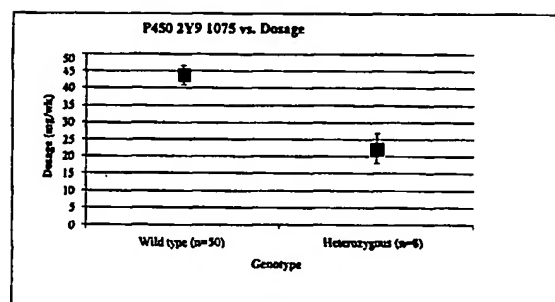
B. vk3294



C. vk4769



D. p1075



Figures 1A-D

2/3

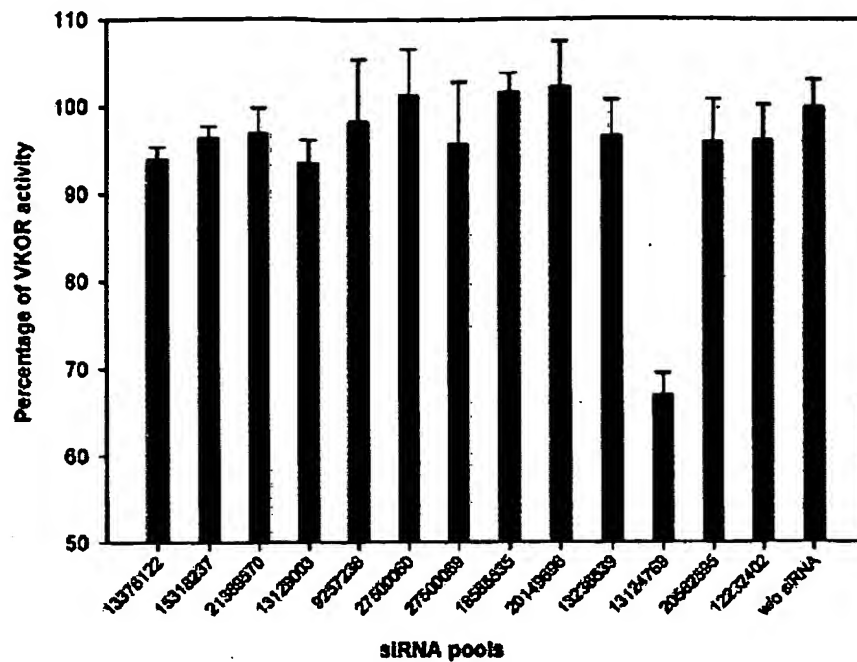


Figure 2

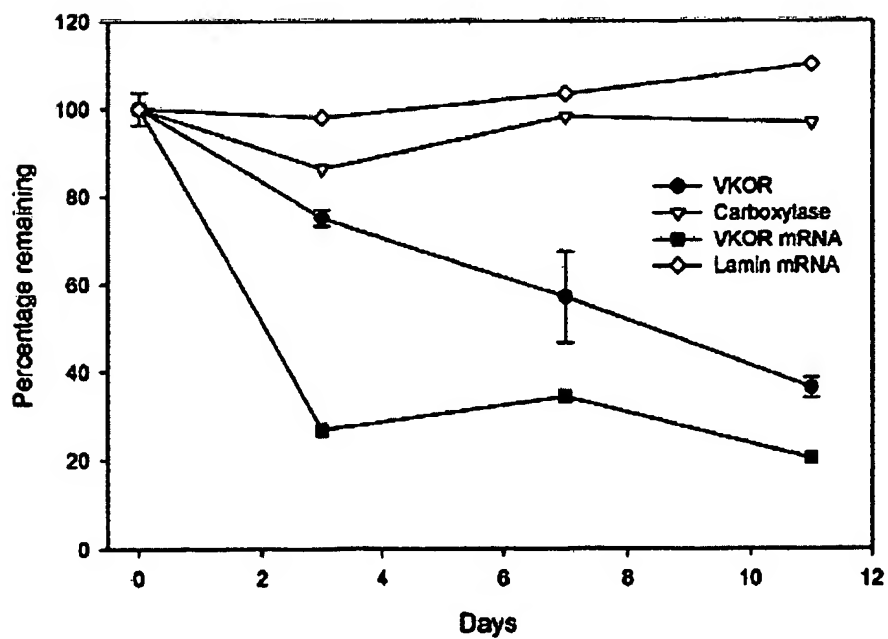


Figure 3

3/3

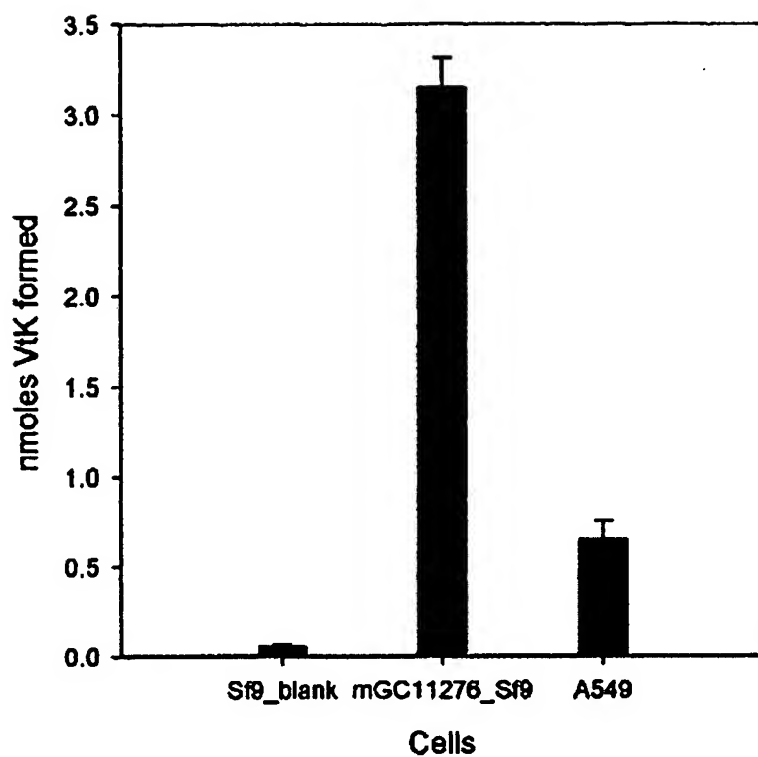


Figure 4

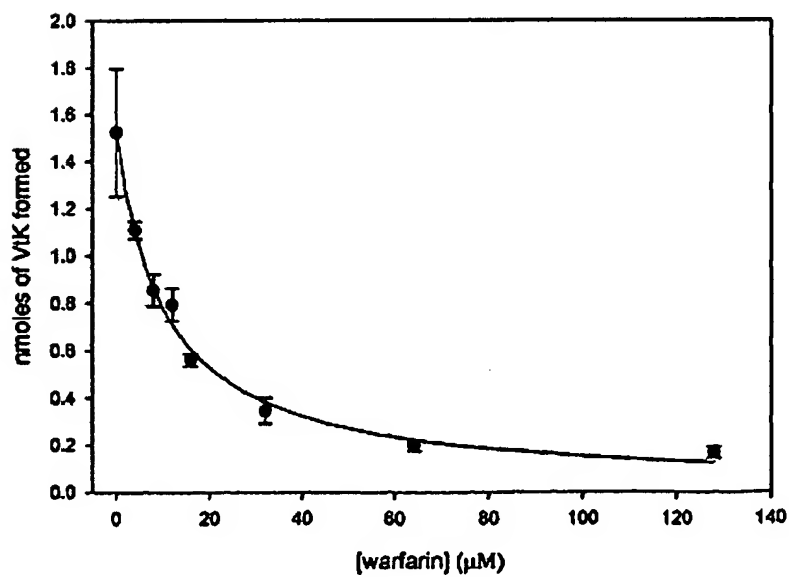


Figure 5

SEQUENCE LISTING

<110> University of North Carolina-Chapel Hill
Stafford, Darrel
Li, Tao

<120> IDENTIFICATION OF THE GENE FOR VITAMIN K EPOXIDE REDUCTASE

<130> 5470.401WO

<150> US 60/505,527
<151> 2003-09-23

<160> 34

<170> PatentIn version 3.2

<210> 1
<211> 21
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide primer

<400> 1
tccaacagca tattcggtg c 21

<210> 2
<211> 21
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide primer

<400> 2
ttcttgacc ttccggaac t 21

<210> 3
<211> 19
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide primer

<400> 3
gaaggtgaag gtcggagtc 19

<210> 4
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide primer

<400> 4
gaagatggtg atgggatttc 20

<210> 5
<211> 21
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide primer

<400> 5
ctaggtgagg ccaagaagca a 21

<210> 6
<211> 21
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide primer

<400> 6
ctgttcctct cagcagactg c 21

<210> 7
<211> 12
<212> PRT
<213> Artificial sequence

<220>
<223> HPC4 tag sequence

<400> 7

Glu Asp Gln Val Asp Pro Arg Leu Ile Asp Gly Lys
1 5 10

<210> 8
<211> 3915
<212> DNA
<213> Homo sapiens

<400> 8
ggttttctcc gcgggcgcct cgggcggaac ctggagataa tgggcagcac ctgggggagc 60
cctggctggg tgccgctcgc tctttgcctg acgggcttag tgctctcgtc ctacgcgctg 120
cacgtgaagg cggcgcgcg cggggaccgg gattaccgcg cgctctgcga cgtgggcacc 180
gccatcagct gttcgcgcgt cttctcctcc aggtgtgcac gggagtggga ggcgtggggc 240
ctcggagcag ggcggccagg atgccagatg attattctgg agtctgggat cgggtgtgcc 300
ggggaacgga cacggggctg gactgctcgc ggggtcgttg cacaggggct gagctaccca 360
gcgatactgg tgttcgaaat aagagtgcga ggcaaggac cagacagtgc tggggactgg 420

gattattccg gggactcgca cgtgaattgg atgccaagga ataacgggtga ccaggaaagg 480
 cggggaggca ggatggcggt agagattgac gatggtctca aggacggcgc gcagggtgaag 540
 gggggtgttg gcgatggctg cgcccaggaa caagggtggcc cgggtctggct gtgcgtgatg 600
 gccaggcggt agcataatga cggaatacag aggaggcgag tgagtggcca gggagctgga 660
 gattctgggg tccaggggcaa agataatctg ccccgactc ccagtctctg atgcaaaacc 720
 gagtgaaccg ttataccagc cttgccattt taagaattac ttaagggccg ggcgcgggtg 780
 cccactcctg taatcccagc actttgggag gccgaggcgg atggatcact tgaagtcagg 840
 agttgaccag cctggccaac atgggtgaaag cctgtctcta ccaaaaatag aaaaattaat 900
 cgggcgctat ggcggtgccc ttaatcccag ctactcgggg gggctaaggc aggagaatcg 960
 cttgaaccgg ggaggcggag gtttcagtga gccgagatcg cgccactgca ctccagcctg 1020
 ggccagagtg agactccgtc tcaaaaaaaaa aaaaaaaaaa aaaaaaaaaa agacttactt 1080
 aagggtctaag atgaaaagca gggcctacgg agtagccacg tccgggcctg gtctggggag 1140
 aggggaggat agggtcagtg acatggaatc ctgacgtggc caaagggtgc cggtgccagg 1200
 agatcatcga cccttgact aggatgggag gtcggggaac agaggatagc ccagggtggct 1260
 tcttggaat cacctttctc gggcagggtc caaggcactg ggttgacagt cctaacctgg 1320
 ttccaccca ccccccct ctgccagggt gggcaggggt ttcgggctgg tggagcatgt 1380
 gctgggacag gacagcatcc tcaatcaatc caacagcata ttcggttgca tcttctacac 1440
 actacagcta ttgttaggtg agtggctccg cccctccct gcccgccccg ccccgcccc 1500
 catccccctt ggtcagctca gcccactcc atgcaatctt ggtgatccac acagctgaca 1560
 gccagctagc tgctcatcac ggagcgtcct gcgggtgggg atgtggggag gtaactaaca 1620
 ggagtctttt aattggttta agtactgtta gaggtgaag ggcccttaa gacatcctag 1680
 gtcccagggt tttttgtttg ttgttgttt gagacagggt ctggctctgt tgcccaaagt 1740
 gaggtctagg atgcccttag tgtgcactgg cgtgatctca gttcatggca acctctgcct 1800
 ccctgccccaa gggatccctc caccttagcc tcccagcag ctggaatcac aggcgtgcac 1860
 cactatgccc agctaatttt tgtttttgtt tttttttggt agagatgggt tctcgccatg 1920
 ttgcccaggc tggctcgaag caatctgtct gcctcagcct cccaaagtgc tggggggatt 1980
 acaggcgtga gctaccatgc cccaccaaca cccagtttt gtggaaaaga tgccgaaatt 2040
 cctttttaag gagaagctga gcatgagcta tcttttgtct catttagtgc tcagcaggaa 2100
 aatttgatc tagtcccata agaacagaga gaggaaccaa gggagtggaa gacgatggcg 2160
 cccaggcct tgctgatgcc atatgccga gatgagacta tccattacca cccttcccag 2220

caggctccca cgctcccttt gagtcaccct tcccagctcc agagaaggca tcaactgaggg 2280
 aggccagca ccatggctct ggctgacaca tgggtcagac ttggccgatt tatttaagaa 2340
 attttattgc tcagaacttt cctccctgg gcaatggcaa gagcttcaga gaccagtccc 2400
 ttggagggga cctgttgaag ccttcttttt tttttttttt aagaaataat cttgctctgt 2460
 tgcccaggct ggagtgcagt ggcacaatca tagctcactg taacctggct caagcgatcc 2520
 tcctgagtag ctaggactat aggcattgca ctgcaccag ctaatttttt tttttttttt 2580
 tttttttttt ttgcgacata gtctcgctct gtcaccaggc tggagtgcag tggcacgatc 2640
 ttggctcact gcaacctctg cctcccggt tcaagcaatt ttctgcctc agcctcctga 2700
 gtagctggga ctacaggcgc gtgtcaccac gccagctaa tttttgtatt ttagtgagg 2760
 acagggttcc accatgttgg ctaggatggt ctcaatctct tgacctggtg atccatccgc 2820
 cttggcctcc caaagtgcct ggattacagg cgtgagctaa cctcaccggg catttttttt 2880
 ttgagacgaa gtcttgctct tgctgcccga gctggaatgt ggtggcatga tctcggtca 2940
 ctgcaacctc cacctcctag gttcaagcga ttctccacct tagcctccc agcagctggg 3000
 attacaggtg cccatcaaca caccggcta atttttgtat ttttattaga gatggggttt 3060
 tgccatgttg gccaggctgc tctcgaaact ctaacctcag gtgatccacc cccattggcc 3120
 tcccaaaata ctgggattac aggcattgag caccgtgccc agctgaattt ctaaattttt 3180
 gatagagatc gggctcttct atgttgccca agctggtctt gaactcctag cctaaagcag 3240
 tcttcccacc tcggcctccc agagtgttg gaatacgtgc gtaagccacc acatctgccc 3300
 tggagcctct tgtttttagag acccttccca gcagctcctg gcatctaggt agtgagtgta 3360
 catcatggag tggtcgggag gtggccagt cctgaagccc acaccggacc ctcttctgcc 3420
 ttgcaggttg cctgcggaca cgctgggcct ctgtcctgat gctgctgagc tccctggtgt 3480
 ctctcgctgg ttctgtctac ctggcctgga tctgttctt cgtgctctat gatttctgca 3540
 ttgtttgtat caccacctat gctatcaacg tgagcctgat gtggctcagt ttccggaagg 3600
 tccaagaacc ccagggaag gctaagaggc actgagccct caaccaagc caggctgacc 3660
 tcatctgctt tgctttggca tgtgagcctt gcctaagggg gcatactctgg gtccctagaa 3720
 ggccttagat gtggggcttc tagattacc cctcctcctg ccataccgc acatgacaat 3780
 ggaccaaagtg tgccacacgc tcgctctttt ttacaccag tgctctgac tctgtcccca 3840
 tgggctggtc tccaaagctc tttccattgc ccagggaggg aagggttctga gcaataaagt 3900
 ttcttagatc aatca 3915

<210> 9

<211> 806

[illegible]

aggaggaggaa gggtctgagc aataaagttt

806

<210> 10
 <211> 163
 <212> PRT
 <213> Homo sapiens

<400> 10

Met Gly Ser Thr Trp Gly Ser Pro Gly Trp Val Arg Leu Ala Leu Cys
 1 5 10 15

Leu Thr Gly Leu Val Leu Ser Leu Tyr Ala Leu His Val Lys Ala Ala
 20 25 30

Arg Ala Arg Asp Arg Asp Tyr Arg Ala Leu Cys Asp Val Gly Thr Ala
 35 40 45

Ile Ser Cys Ser Arg Val Phe Ser Ser Arg Trp Gly Arg Gly Phe Gly
 50 55 60

Leu Val Glu His Val Leu Gly Gln Asp Ser Ile Leu Asn Gln Ser Asn
 65 70 75 80

Ser Ile Phe Gly Cys Ile Phe Tyr Thr Leu Gln Leu Leu Leu Gly Cys
 85 90 95

Leu Arg Thr Arg Trp Ala Ser Val Leu Met Leu Leu Ser Ser Leu Val
 100 105 110

Ser Leu Ala Gly Ser Val Tyr Leu Ala Trp Ile Leu Phe Phe Val Leu
 115 120 125

Tyr Asp Phe Cys Ile Val Cys Ile Thr Thr Tyr Ala Ile Asn Val Ser
 130 135 140

Leu Met Trp Leu Ser Phe Arg Lys Val Gln Glu Pro Gln Gly Lys Ala
 145 150 155 160

Lys Arg His

<210> 11
 <211> 5915
 <212> DNA
 <213> Homo sapiens

<400> 11

caccatcaga tgggacgtct gtgaaggaga gacctcatct ggcccacagc ttggaaagga

60

gagactgact gttgagttga tgcaagctca ggtgttgcca ggcgggcgcc atgatagtag 120
 agaggtagg atactgtcaa ggggtgtgtg ggccaaagga gtggttctgt gaatgtatgg 180
 gagaaaggga gaccgaccac caggaagcac tggtagggca ggaccggga ggatgggagg 240
 ctgcagcccg aatggtgcct gaaatagttt caggggaaat gcttggttcc cgaatcggat 300
 cgccgtattc gctggatccc ctgatccgct ggtctctagg tcccggatgc tgcaattctt 360
 acaacaggac ttggcatagg gtaagcgcaa atgctgttaa ccacactaac acactttttt 420
 ttttcttttt tttttttgag acagagtctc actctgtcgg cctggctgga gtgcagtggc 480
 acgatctcgg ctactgcaa cctccggctc cccggctcaa gcaattctcc tgcctcagcc 540
 tcccagtag ctgggattac aggcattgtc caccacgccc ggctaatttt tgtattttta 600
 gttgagatgg ggtttcacca tgttggcgag gctggtcttg aactcctgac ctgaggaat 660
 ccgccagcct cggcctccca aagtgtctgg attacaagcg tgagccaccg tgcccgccca 720
 acagttttta aatctgtgga gacttcattt cccttgatgc cttgcagccg cgccgactac 780
 aactcccatc atgcctggca gccgctgggg ccgcgattcc gcacgtccct taccgcttc 840
 actagtcccg gcattcttcg ctgttttcct aactcgcccg cttgactagc gccctggaac 900
 agccatttgg gtcgtggagt gcgagcacgg ccggccaatc gccgagtcag agggccagga 960
 ggggcgcggc cattcgccgc ccggccctg ctccgtggct ggttttctcc gcgggcgcct 1020
 cgggcggaac ctggagataa tgggcagcac ctgggggagc cctggctggg tgcggctcgc 1080
 tctttgcctg acgggcttag tgctctcgct ctacgcgctg cacgtgaagg cggcgcgcg 1140
 ccgggaccgg gattaccgcg cgctctgcga cgtgggcacc gccatcagct gttcgcgct 1200
 cttctctccc aggtgtgcac gggagtggga ggcgtggggc ctcgagcag ggcggccagg 1260
 atgccagatg attattctgg agtctgggat cgggtgtccc ggggaacgga cacggggctg 1320
 gactgctcgc ggggtcgtg cacaggggct gagctacca gcgatactgg tgttcgaaat 1380
 aagagtgcga ggcaaggac cagacagtgc tggggactgg gattattccg gggactcgca 1440
 cgtgaattgg atgccaagga ataacggtga ccaggaaagg cggggaggca ggatggcggt 1500
 agagattgac gatggtctca aggacggcgc gcaggtgaag gggggtgttg gcgatggctg 1560
 cgcccaggaa caaggtggcc cggctctggct gtgcgtgatg gccaggcggt agcataatga 1620
 cggaatacag aggaggcgag tgagtggcca gggagctgga gattctgggg tccagggcaa 1680
 agataatctg ccccgactc ccagtctctg atgcaaaacc gagtgaaccg ttataaccagc 1740
 cttgccattt taagaattac ttaagggcgg ggcgcggtgg cccactcctg taatccagc 1800
 actttgggag gccgaggcgg atggatcact tgaagtcagg agttgaccag cctggccaac 1860

atgggtgaaag cctgtctcta ccaaaaatag aaaaattaat cgggcgctat ggcgggtgcc 1920
 ttaatcccag ctactcgggg gggctaaggc aggagaatcg cttgaacccg ggaggcggag 1980
 gtttcagtga gccgagatcg cgccactgca ctccagcctg ggccagagtg agactccgtc 2040
 tcaaaaaaaaa aaaaaaaaaa aaaaaaaaaag agacttactt aaggtctaag atgaaaagca 2100
 gggcctacgg agtagccacg tccgggcctg gtctggggag aggggaggat agggtcagtg 2160
 acatggaatc ctgacgtggc caaagggtgcc cggtgccagg agatcatcga cccttggaact 2220
 aggatgggag gtcggggaac agaggatagc ccagggtggct tcttggaat cacccttctc 2280
 gggcagggtc caaggcactg ggttgacagt cctaacctgg ttccacccca cccacccct 2340
 ctgccagggtg gggcaggggt ttccggctgg tggagcatgt gctgggacag gacagcatcc 2400
 tcaatcaatc caacagcata ttccggttgc tcttctacac actacagcta ttgttaggtg 2460
 agtggctccg cccctccct gcccgcccg ccccgccct catccccctt ggtcagctca 2520
 gcccactcc atgcaatctt ggtgatccac acagctgaca gccagctagc tgctcatcac 2580
 ggagcgtcct gcgggtgggg atgtggggag gtaactaaca ggagtctttt aattggttta 2640
 agtactgtta gaggtgaag ggccttaaa gacatcctag gtcccaggt tttttgtttg 2700
 ttgttgtttt gagacagggt ctggctctgt tgcccaaagt gaggtctagg atgcccttag 2760
 tgtgactgg cgtgatctca gttcatggca acctctgcct ccctgcccga gggatcctcc 2820
 caccttagcc tccaagcag ctggaatcac aggcgtgcac cactatgcc agctaatttt 2880
 tgtttttgtt ttttttttgt agagatggtg tctcgccatg ttgccaggc tggctcctcaag 2940
 caatctgtct gcctcagcct cccaaagtgc tggggggatt acaggcgtga gctaccatgc 3000
 cccaccaaca cccagtttt gtggaaaaga tgccgaaatt cttttttaag gagaagctga 3060
 gcatgagcta tcttttgtct catttagtgc tcagcaggaa aatttgtatc tagtcccata 3120
 agaacagaga gaggaaccaa gggagtggaa gacgatggcg cccaggcct tgctgatgcc 3180
 atatgccgga gatgagacta tccattacca ccctcccag caggctcca cgctcccttt 3240
 gagtcaccct tccagctcc agagaaggca tctactgagg agggccagca ccatggctct 3300
 ggctgacaca tggttcagac ttggccgatt tatttaagaa attttattgc tcagaacttt 3360
 ccctccctgg gcaatggcaa gagcttcaga gaccagtccc ttggagggga cctgttgaag 3420
 ccttcttttt tttttttttt aagaaataat cttgctctgt tgcccaggct ggagtgcagt 3480
 ggcacaatca tagctcactg taacctggct caagcgatcc tcctgagtag ctaggactat 3540
 aggcagtgtc ctgcacccag ctaatttttt tttttttttt tttttttttt ttgcgacata 3600
 gtctcgctct gtcaccaggc tggagtgcag tggcacgac ttggctcact gcaacctctg 3660
 cctcccggt tcaagcaatt ttctgcctc agcctcctga gtagctggga ctacaggcgc 3720

gtgtcaccac gccagctaa tttttgtatt tttagtggag acagggtttc accatgttgg 3780
 ctaggatggg ctcaatctct tgacctggg atccatccgc cttggcctcc caaagtgcta 3840
 ggattacagg cgtgagtcaa cctcaccggg catttttttt ttgagacgaa gtcttgctct 3900
 tgctgcccac gctggaatgt ggtggcatga tctcggtcca ctgcaacctc cacctcctag 3960
 gttcaagcga ttctccacct tagcctcccc agcagctggg attacagggtg cccatcaaca 4020
 caccgggcta atttttgtat ttttattaga gatgggggtt tgccatgttg gccaggctgc 4080
 tctcgaactc ctaacctcag gtgatccacc ccatttggcc tcccaaaata ctgggattac 4140
 aggcattgagc caccgtgccc agctgaatct ctaaattttt gatagagatc ggggtctttct 4200
 atgttgccca agctggctct gaactcctag cctaaagcag tcttcccacc tcggcctccc 4260
 agagtgtttg gaatacgtgc gtaagccacc acatctgccc tggagcctct tgttttagag 4320
 acccttccca gcagctcctg gcattctagg agtgacgtga catcatggag tgttcgggag 4380
 gtggccagtg cctgaagccc acaccggacc ctcttctgcc ttgcaggttg cctgcggaca 4440
 cgctgggcct ctgtctgat gctgctgagc tccctgggtg ctctcgctgg ttctgtctac 4500
 ctggcctgga tctgttctt cgtgctctat gatttctgca ttgtttgtat caccacctat 4560
 gctatcaacg tgagcctgat gtggctcagt ttccggaagg tccaagaacc ccagggaag 4620
 gctaagaggc actgagccct caaccaagc caggctgacc tcatctgctt tgctttggca 4680
 tgtgagcctt gcctaagggg gcatactctg gtccctagaa ggccctagat gtggggcttc 4740
 tagattaccc cctcctcctg ccataccgc acatgacaat ggaccaaagtg tgccacacgc 4800
 tcgctctttt ttacaccag tgcctctgac tctgtcccca tgggctggtc tccaaagctc 4860
 tttccattgc ccaggaggag aaggttctga gcaataaagt ttcttagatc aatcagccaa 4920
 gtctgaacca tgtgtctgcc atggactgtg gtgctgggcc tccctcggtg ttgccttctc 4980
 tggagctggg aagggtgagt cagagggaga gtggagggcc tgctgggaag ggtggttatg 5040
 ggtagtctca tctccagtgt gtggagtcag caaggcctgg ggcaccattg gccccaccc 5100
 ccaggaaaca ggctggcagc tcgctcctgc tgcccacagg agccaggcct cctctcctgg 5160
 gaaggctgag cacacacctg gaagggcagg ctgcccttct ggttctgtaa atgcttgctg 5220
 ggaagtctct ccttgagttt aactttaacc cctccagtgt ccttatcgac cattccaagc 5280
 cagtattggt agccttggag ggtcagggcc aggttgtaga ggtttttgtt ttgcctatta 5340
 tgccctgacc acttacctac atgccaagca ctgtttaaga acttggtgtg gcagggtgca 5400
 gtggctcaca cctgtaatec ctgtactttg ggaggccaag gcaggaggat cacttgaggc 5460
 caggagttcc agaccagcct gggcaaaata gtgagacccc tgtctctaca aaaaaaaaaa 5520

aaaaaaaaaa ttagccaggc atggtggtgt atgtacctat agtcccaact aatcggaag 5580
 ctggcgggaa gactgcttga gccagaagg ttgaggctgc agtgagccat gatcactgca 5640
 ctccagcctg agcaacagag caagaccgtc tccaaaaaaa aacaaaaaac aaaaaaaac 5700
 ttgtgttaac gtgttaaact cgtttaatct ttacagtgat ttatgagggtg ggtactatta 5760
 ttatccctat cttgatgata gggacagagt ggctaattag tatgcctgag atcacacagc 5820
 tactgcagga ggctctcagg atttgaatcc acctgggtcca tctgggtcca gcatctatat 5880
 gctttttttt ttgttggttt gtttttgaga cggac 5915

<210> 12

<211> 5915

<212> DNA

<213> Homo sapiens

<400> 12

caccatcaga tgggacgtct gtgaaggaga gacctcatct ggcccacagc ttggaagga 60
 gagactgact gttgagttga tgcaagctca ggtgttgcca ggcgggcgcc atgatagtag 120
 agaggttagg atactgtcaa ggggtgtgtgt ggccaaagga gtggttctgt gaatgtatgg 180
 gagaaagga gaccgaccac caggaagcac tggtgaggca ggacccggga ggatgggagg 240
 ctgcagcccg aatgggtgct gaaatagttt caggggaaat gcttggttcc cgaatcggat 300
 cgccgtattc gctggatccc ctgatccgct ggtctctagg tcccggatgc tgcaattctt 360
 acaacaggac ttggcatagg gtaagcgcaa atgctgttaa ccacactaac acactttttt 420
 ttttcttttt tttttttgag acagagtctc actctgtcgg cctggctgga gtgcagtggc 480
 acgatctcgg ctactgcaa cctccggctc cccggctcaa gcaattctcc tgcctcagcc 540
 tcccagtag ctgggattac aggcattgtc caccacgccc ggctaatttt tgtattttta 600
 gttgagatgg ggtttcacca tgttggcgag gctggtcttg aactcctgac ctgaggtaat 660
 ccgccagcct cggcctccca aagtgtggg attacaagcg tgagccaccg tgcccggcca 720
 acagttttta aatctgtgga gacttcattt cccttgatgc cttgcagccg cgccgactac 780
 aactcccatc atgcctggca gccgtgggg ccgcgattcc gcacgtccct taccgcttc 840
 actagtcccg gcattcttcg ctgttttctt aactcgcccg cttgactagc gccctggaac 900
 agccatttgg gtcgtggagt gcgagcacgg ccggccaatc gccgagtcag agggccagga 960
 ggggcgcggc cattcgccgc ccggcccttg ctccgtggct ggttttctcc gcgggcgcct 1020
 cgggcggaac ctggagataa tgggcagcac ctgggggagc cctggctggg tgcggtctgc 1080
 tctttgcctg acgggcttag tgctctcgtc ctacgcgtcg cacgtgaagg cggcgcgccg 1140
 ccgggaccgg gattaccgcg cgctctcgca cgtgggcacc gccatcagct gttcgcgcgt 1200

cttctcctcc aggtgtgcac gggagtggga ggcgtggggc ctcgagcag ggcggccagg 1260
 atgccagatg attattctgg agtctgggat cgggtgtccc ggggaacgga cacggggctg 1320
 gactgctcgc ggggtcgttg cacaggggct gagctacca gcgatactgg tgttcgaaat 1380
 aagagtgcga ggcaaggac cagacagtgc tggggactgg gattattccg gggactcgca 1440
 cgtgaattgg atgccaagga ataacggtga ccaggaaagg cggggaggca ggatggcggt 1500
 agagattgac gatggtctca aggacggcgc gcaggtaag gggggtgttg gcgatggctg 1560
 cgcccaggaa caaggtggcc cggctctggct gtgcgtgatg gccaggcgtt agcataatga 1620
 cggaatacag aggagggcag tgagtggcca gggagctgga gattctgggg tccagggcaa 1680
 agataatctg cccccactc ccagtctctg atgcaaaacc gagtgaaccg ttataaccagc 1740
 cttgccatth taagaattac ttaagggccg ggcgcgggtg cccactcctg taatcccagc 1800
 actttgggag gccgagggcg atggatcact tgaagtcagg agttgaccag cctggccaac 1860
 atggtgaaag cctgtctcta ccaaaaatag aaaaattaat cgggcgctat ggcgggtgcc 1920
 ttaatcccag ctactcgggg gggctaaggc aggagaatcg cttgaaccg ggagggcgag 1980
 gtttcagtga gccgagatcg cgccactgca ctccagcctg ggccagagt agactccgtc 2040
 tcaaaaaaaaa aaaaaaaaaa aaaaaaaaaa agacttactt aaggtctaag atgaaaagca 2100
 gggcctacgg agtagccacg tccgggcctg gtctggggag aggggaggat agggtcagtg 2160
 acatggaatc ctgacgtggc caaagggtgc cgggtgccagg agatcatcga cccttgact 2220
 aggatgggag gtcggggaac agaggatagc ccagggtggct tcttggaat cacctttctc 2280
 gggcagggtc caaggcactg ggttgacagt cctaacctg tccacccca cccaccct 2340
 ctgccaggtg gggcaggggt ttcgggctg tggagcatgt gctgggacag gacagcatcc 2400
 tcaatcaatc caacagcata ttcggttgca tcttctacac actacagcta ttgttaggtg 2460
 agtggctccg cccctccct gcccgccccg cccgcccct catccccctt ggtcagctca 2520
 gcccactcc atgcaatctt ggtgatccac acagctgaca gccagctagc tgctcatcac 2580
 cgagcgtcct ggggtgggg atgtggggag gtaactaaca ggagtctttt aattggttta 2640
 agtactgtta gaggtgaag ggcccttaaa gacatcctag gtccccagg tttttgttg 2700
 ttgttgtttt gagacagggt ctggctctgt tgcccaaagt gaggtctagg atgcccttag 2760
 tgtgactgg cgtgatctca gttcatggca acctctgcct ccctgcccga gggatcctcc 2820
 caccttagcc toccaagcag ctggaatcac aggcgtgcac cactatgcc agctaatttt 2880
 tgtttttgtt tttttttgt agagatgggt tctcgccatg ttgccaggc tgggtctcaag 2940
 caatctgtct gcctcagcct cccaaagtgc tggggggatt acaggcgtga gctaccatgc 3000
 cccaccaaca cccagtttt gtggaaaaga tgccgaaatt cttttttaag gagaagctga 3060

gcatgagcta tcttttgtct catttagtgc tcagcaggaa aatttgtatc tagtcccata 3120
agaacagaga gaggaaccaa gggagtggaa gacgatggcg cccaggcct tgctgatgcc 3180
atatgccgga gatgagacta tccattacca cccttcccag caggctccca cgctcccttt 3240
gagtcaccct tcccagctcc agagaaggca tcaactgagg aggcccagca ccatggctct 3300
ggctgacaca tggttcagac ttggccgatt tatttaagaa attttattgc tcagaacttt 3360
ccctccctgg gcaatggcaa gagcttcaga gaccagtccc ttggagggga cctgttgaag 3420
ccttcttttt tttttttttt aagaaataat cttgctctgt tgcccaggct ggagtgcagt 3480
ggcacaatca tagctcactg taacctggct caagcgatcc tcctgagtag ctaggactat 3540
aggcatgtca ctgcaccag ctaatttttt tttttttttt tttttttttt ttgcgacata 3600
gtctcgctct gtcaccaggc tggagtgcag tggcacgatc ttggctcact gcaacctctg 3660
cctcccggt tcaagcaatt ttctgcctc agcctcctga gtagctggga ctacaggcgc 3720
gtgtcaccac gccagctaa tttttgtatt tttagtggag acagggtttc accatgttgg 3780
ctaggatggt ctcaatctct tgacctggtg atccatccgc cttggcctcc caaagtgcta 3840
ggattacagg cgtgagtcaa cctcaccggg catttttttt ttgagacgaa gtcttgctct 3900
tgctgcccga gctggaatgt ggtggcatga tctcggtca ctgcaacctc cacctcctag 3960
gttcaagcga ttctccacct tagcctcccc agcagctggg attacagggtg cccatcaaca 4020
caccggcta atttttgtat ttttattaga gatggggttt tgccatgttg gccaggctgc 4080
tctcgaactc ctaacctcag gtgatccacc cccattggcc tcccaaaata ctgggattac 4140
aggcatgagc caccgtgccc agctgaatth ctaaattttt gatagagatc gggctcttct 4200
atgttgccca agctggtctt gaactcctag cctaaagcag tcttcccacc tcggcctccc 4260
agagtgtttg gaatacgtgc gtaagccacc acatctgccc tggagcctct tgttttagag 4320
accttccca gcagctcctg gcatctaggt agtgagtgat catcatggag tgttcgggag 4380
gtggccagt cctgaagccc acaccggacc ctcttctgcc ttgcaggttg cctgcggaca 4440
cgctgggcct ctgtcctgat gctgctgagc tccctggtgt ctctcgctgg ttctgtctac 4500
ctggcctgga tctgttctt cgtgctctat gatttctgca ttgtttgtat caccacctat 4560
gctatcaacg tgagcctgat gtggctcagt ttccggaagg tccaagaacc ccagggcaag 4620
gctaagaggc actgagccct caaccaagc caggctgacc tcatctgctt tgctttggca 4680
tgtgagcctt gcctaagggg gcatatctgg gtccctagaa ggccctagat gtggggcttc 4740
tagattaccc cctcctcctg ccataccgc acatgacaat ggaccaaag tgccacacgc 4800
tcgctctttt ttacaccag tgcctctgac tctgtccca tgggctggtc tccaaagctc 4860

```

tttccattgc ccagggaggg aaggttctga gcaataaagt ttcttagatc aatcagccaa 4920
gtctgaacca tgtgtctgcc atggactgtg gtgctggggc tccctcgggtg ttgccttctc 4980
tgtagctggg aagggtgagt cagagggaga gtggagggcc tgctgggaag ggtggttatg 5040
ggtagtctca tctccagtgt gtggagtcag caaggcctgg ggcaccattg gccccacccc 5100
ccaggaaaca ggctggcagc tcgctcctgc tgcccacagg agccaggcct cctctcctgg 5160
gaaggctgag cacacacctg gaagggcagg ctgcccttct ggttctgtaa atgcttgctg 5220
ggaagttctt ccttgagttt aactttaacc cctccagtgt ccttatcgac cattccaagc 5280
cagtattggt agccttgagg ggtcagggcc aggttgtgaa ggtttttgtt ttgcctatta 5340
tgccctgacc acttacctac atgccaagca ctgtttaaga acttgtgttg gcagggtgca 5400
gtggctcaca cctgtaatcc ctgtactttg ggaggccaag gcaggaggat cacttgaggc 5460
caggagtcc agaccagcct gggcaaaata gtgagacccc tgtctctaca aaaaaaaaaa 5520
aaaaaaaaa ttagccaggc atgggtggtg atgtacctat agtcccaact aatcggaag 5580
ctggcgggaa gactgcttga gcccagaagg ttgaggctgc agtgagccat gatcactgca 5640
ctccagcctg agcaacagag caagaccgtc tccaaaaaaa aacaaaaaac aaaaaaaac 5700
ttgtgttaac gtgttaaact cgtttaactt ttacagtgat ttatgagggtg ggtactatta 5760
ttatccctat cttgatgata gggacagagt ggctaattag tatgcctgag atcacacagc 5820
tactgcagga ggctctcagg atttgaatcc acctgggtcca tctgggtcca gcatctatat 5880
gctttttttt ttgttggttt gtttttgaga cggac 5915

```

<210> 13
<211> 5915
<212> DNA
<213> Homo sapiens

```

<400> 13
caccatcaga tgggacgtct gtgaaggaga gacctcatct ggcccacagc ttgaaagga 60
gagactgact gttgagttga tgcaagctca ggtgttgcca ggcgggcgcc atgatatag 120
agaggtagg atactgtcaa ggggtgtgtg ggccaaagga gtggttctgt gaatgtatg 180
gagaaagga gaccgaccac caggaagcac tggtaggca ggaccggga ggatgggagg 240
ctgcagcccg aatgggtgcct gaaatagttt caggggaaat gcttggttcc cgaatcggat 300
cgccgtattc gctggatccc ctgatccgct ggtctctagg tcccggatgc tgcaattctt 360
acaacaggac ttggcatagg gtaagcgcaa atgctgttaa ccacactaac acactttttt 420
ttttcttttt tttttttgag acagagtctc actctgtcgg cctgggtgga gtgcagtggc 480
acgatctcgg ctactgcaa cctccggctc cccgggtcaa gcaattctcc tgccctcagc 540

```

tcccagtag ctgggattac aggcattgtgc caccacgccc ggctaatttt tgtattttta	600
gttgagatgg ggtttcacca tgttggcgag gctggctttg aactcctgac ctcaggtaat	660
ccgccagcct cggcctccca aagtgtctggg attacaagcg tgagccaccg tgcccggcca	720
acagttttta aatctgtgga gacttcattt cccttgatgc cttgcagccg cgccgactac	780
aactcccatc atgcctggca gccgctgggg ccgcgattcc gcacgtccct taccgcttc	840
actagtcccg gcattcttcg ctgttttctt aactcgcccg cttgactagc gccctggaac	900
agccatttgg gtcgtggagt gcgagcacgg ccggccaatc gccgagtcag agggccagga	960
ggggcgcggc cattcgccgc ccggccccctg ctccgtggct ggttttctcc gggggcgct	1020
cgggcggaac ctggagataa tgggcagcac ctgggggagc cctggctggg tgccgctcgc	1080
tctttgcctg acgggcttag tgctctcgt ctacgcgtg cacgtgaagg cggcgcgcg	1140
ccgggaccgg gattaccgcg cgctctcgca cgtgggcacc gccatcagct gttcgcgct	1200
cttctctccc aggtgtgcac gggagtggga ggcgtggggc ctcgagcag ggcggccagg	1260
atgccagatg attattctgg agtctgggat cgggtgtccc ggggaacgga cacggggctg	1320
gactgctcgc ggggtcgtg cacaggggct gagctacca gcgatactgg tgttcgaaat	1380
aagagtgcga ggcaaggac cagacagtgc tggggactgg gattattccg gggactcgca	1440
cgtgaattgg atgccaagga ataacgtga ccaggaaagg cggggaggca ggatggcgg	1500
agagattgac gatggtctca aggacggcg gcaggtgaag gggggtgtg gcgatggctg	1560
cgcccaggaa caagggtggc cggctctggct gtgcgtgatg gccaggcggt agcataatga	1620
cggaatacag aggaggcgag tgagtggcca gggagctgga gattctgggg tccagggcaa	1680
agataatctg ccccgactc ccagtctctg atgcaaaacc gagtgaaccg ttataccagc	1740
cttgccattt taagaattac ttaagggccg ggcgcggtgg ccactcctg taatcccagc	1800
actttgggag gccgaggcgg atggatcact tgaagtcagg agttgaccag cctggccaac	1860
atgggtgaaag cctgtctcta ccaaaaatag aaaaattaat cgggcgctat ggcgggtgcc	1920
ttaatcccag ctactcgggg gggctaaggc aggagaatcg cttgaaccgg ggaggcggag	1980
gtttcagtga gccgagatcg cgccactgca ctccagcctg ggccagagtg agactccgtc	2040
tcaaaaaaaaa aaaaaaaaaa aaaaaaaaaa agacttactt aaggtctaag atgaaaagca	2100
gggcctacgg agtagccacg tccgggcctg gtctggggag aggggaggat agggtcagt	2160
acatggaatc ctgacgtggc caaagggtgc cgggtgccagg agatcatcga cccttgga	2220
aggatgggag gtcggggaac agaggatagc ccaggtggct tcttggaat cacctttctc	2280
gggcagggtc caaggcactg ggttgacagt cctaacctgg ttccaccca cccacccct	2340
ctgccagggtg gggcaggggt ttcgggctgg tggagcatgt gctgggacag gacagcatcc	2400

tcaatcaatc caacagcata ttcggttgca tcttctacac actacagcta ttgttaggtg	2460
agtggctccg cccctccct gcccgccccg ccccgccct catccccctt ggtcagctca	2520
gcccactcc atgcaatctt ggtgatccac acagctgaca gccagctagc tgctcatcac	2580
ggagcgtcct gcgggtggg atgtggggag gtaactaaca ggagtctttt aattggttta	2640
agtactgtta gaggctgaag ggcccttaaa gacatcctag gtccccaggt tttttgtttg	2700
ttgttgtttt gagacagggt ctggctctgt tgcccaaagt gaggtctagg atgcccttag	2760
tgtgcactgg cgtgatctca gttcatggca acctctgcct ccctgcccaa gggatccctcc	2820
caccttagcc tccaagcag ctggaatcac aggcgtgcac cactatgccc agctaatttt	2880
tgtttttgtt ttttttgggt agagatgggt tctcgccatg ttgcccaggc tgggtctcaag	2940
caatctgtct gcctcagcct cccaaagtgc tggggggatt acaggcgtga gctaccatgc	3000
cccaccaaca cccagtttt gtggaaaaga tgccgaaatt cttttttaag gagaagctga	3060
gcatgagcta tcttttgtct catttagtgc tcagcaggaa aatttgtatc tagtcccata	3120
agaacagaga gaggaacca gggagtggaa gacgatggcg cccaggcct tgctgatgcc	3180
atatgccgga gatgagacta tccattacca cccttcccag caggtccca cgctcccttt	3240
gagtcaccct tcccagctcc agagaaggca tcaactgagg aggccagca ccacggtcct	3300
ggctgacaca tggttcagac ttggccgatt tatttaagaa attttattgc tcagaacttt	3360
ccctccctgg gcaatggcaa gagcttcaga gaccagtccc ttggagggga cctgttgaag	3420
ccttcttttt tttttttttt aagaaataat cttgctctgt tgcccaggct ggagtgcagt	3480
ggcacaatca tagtctactg taacctggct caagcgatcc tccctgagtag ctaggactat	3540
aggcatgtca ctgcaccag ctaatttttt tttttttttt tttttttttt ttgcgacata	3600
gtctcgctct gtcaccaggc tggagtgcag tggcacgac ttggctcact gcaacctctg	3660
cctcccggt tcaagcaatt ttcctgcctc agcctcctga gtagctggga ctacaggcgc	3720
gtgtcaccac gccagctaa tttttgtatt tttagtggag acagggtttc accatgttgg	3780
ctaggatggt ctcaatctct tgacctggtg atccatccgc cttggcctcc caaagtgcta	3840
ggattacagg cgtgagtcaa cctcaccggg catttttttt ttgagacgaa gtcttgcctc	3900
tgctgcccga gctggaatgt ggtggcatga tctcggtcga ctgcaacctc cacctcctag	3960
gttcaagcga ttctccacct tagcctcccc agcagctggg attacagggt cccatcaaca	4020
caccgggcta atttttgtat ttttattaga gatgggggtt tgccatgttg gccaggctgc	4080
tctcgaactc ctaacctcag gtgatccacc ccattggcc tcccaaaata ctgggattac	4140
aggcatgagc caccgtgccc agctgaattt ctaaattttt gatagagatc gggctcttct	4200

atgttgccca agctggtctt gaactcctag cctaaagcag tcttcccacc tcggcctccc 4260
agagtgtttg gaatacgtgc gtaagccacc acatctgccc tggagcctct tgttttagag 4320
acccttccca gcagctcctg gcatctaggt agtgcagtga catcatggag tggtcgggag 4380
gtggccagtg cctgaagccc acaccggacc ctcttctgcc ttgcagggtg cctgcggaca 4440
cgctgggcct ctgtcctgat gctgctgagc tccctgggtg ctctcgctgg ttctgtctac 4500
ctggcctgga tctgttctt cgtgctctat gatttctgca ttgtttgtat caccacctat 4560
gctatcaacg tgagcctgat gtggctcagt ttccggaagg tccaagaacc ccagggcaag 4620
gctaagaggc actgagccct caacccaagc caggtctgacc tcatctgctt tgctttggca 4680
tgtgagcctt gcctaagggg gcatatctgg gtccctagaa ggccctagat gtggggcttc 4740
tagattaccc cctcctcctg ccatacccg ccatgacaat ggaccaaag tgccacacgc 4800
tcgctctttt ttacaccag tgcctctgac tctgtcccca tgggctgggc tccaaagctc 4860
tttccattgc ccaggaggag aaggttctga gcaataaagt ttcttagatc aatcagccaa 4920
gtctgaacca tgtgtctgcc atggactgtg gtgctgggccc tccctcggtg ttgccttctc 4980
tgagctggg aagggtgagt cagagggaga gtggagggcc tgctgggaag ggtggttatg 5040
ggtagtctca tctccagtgt gtggagtcat caaggcctgg ggcaccattg gccccaccc 5100
ccaggaaaca ggctggcagc tcgctcctgc tgccacagg agccaggcct cctctcctgg 5160
gaaggctgag cacacacctg gaagggcagg ctgcccttct ggttctgtaa atgcttctg 5220
ggaagtctt ccttgagttt aactttaacc cctccagtgt ccttatcgac cattccaagc 5280
cagtattggt agccttgag ggtcagggcc aggttgtgaa ggtttttgtt ttgcctatta 5340
tgccctgacc acttacctac atgccaagca ctgtttaaga acttgtgttg gcagggtgca 5400
gtggctcaca cctgtaatcc ctgtactttg ggaggccaag gcaggaggat cacttgaggc 5460
caggagtctc agaccagcct gggcaaaata gtgagacccc tgtctctaca aaaaaaaaaa 5520
aaaaaaaaa ttagccaggc atggtggtgt atgtacctat agtcccaact aatcggaag 5580
ctggcgggaa gactgcttga gcccagaagg ttgaggctgc agtgagccat gatcactgca 5640
ctccagcctg agcaacagag caagaccgtc tccaaaaaaa aacaaaaaac aaaaaaaac 5700
ttgtgttaac gtgttaaac ctgttaact ttacagtgat ttatgagggtg ggtactatta 5760
ttatccctat cttgatgata gggacagagt ggctaattag tatgcctgag atcacacagc 5820
tactgcagga ggctctcagg atttgaatcc acctgggtcca tctgggtcca gcatctatat 5880
gctttttttt ttgttggttt gtttttgaga cggac 5915

<210> 14

<211> 5915

<212> DNA

<213> Homo sapiens

<400> 14

caccatcaga tgggacgtct gtgaaggaga gacctcatct ggcccacagc ttggaaagga	60
gagactgact gttgagttga tgcaagctca ggtgttgcca ggcgggcgcc atgatagtag	120
agaggttagg atactgtcaa ggggtgtgtg ggccaaagga gtggttctgt gaatgtatgg	180
gagaaaggga gaccgaccac caggaagcac tggtaggca ggacccggga ggatgggagg	240
ctgcagcccg aatgggtgcct gaaatagttt caggggaaat gcttggttcc cgaatcggat	300
cgccgtattc gctggatccc ctgatccgct ggtctctagg tcccggatgc tgcaattcct	360
acaacaggac ttggcatagg gtaagcgcaa atgctgttaa ccacactaac acactttttt	420
ttttcttttt tttttttgag acagagtctc actctgtcgg cctggctgga gtgcagtggc	480
acgatctcgg ctactgcaa cctccggctc cccggctcaa gcaattctcc tgcctcagcc	540
tcccgagtag ctgggattac aggcagtgc caccacgccc ggctaatttt tgtattttta	600
gttgagatgg ggtttcacca tgttggcgag gctggtcttg aactcctgac ctgaggaat	660
ccgccagcct cggcctccca aagtgtggg attacaagcg tgagccaccg tgcccggcca	720
acagttttta aatctgtgga gacttcattt cccttgatgc cttgcagccg cgccgactac	780
aactcccatc atgcctggca gccgctgggg ccgcgattcc gcacgtccct taccgcttc	840
actagtcccg gcattcttcg ctgttttcct aactcgcccc cttgactagc gccctggaac	900
agccatttgg gtcgtggagt gcgagcacgg ccggccaatc gccgagtcag agggccagga	960
ggggcgcggc cattcgccgc ccggccccctg ctccgtggct ggttttctcc gcgggcgccct	1020
cgggcggaac ctggagataa tgggcagcac ctgggggagc cctggctggg tgcggctcgc	1080
tctttgcctg acgggcttag tgctctcgt ctacgcgctg cacgtgaagg cggcgcgcg	1140
ccgggaccgg gattaccgcg cgctctgcga cgtgggcacc gccatcagct gttcgcgcgt	1200
cttctcctcc aggtgtgcac gggagtggga ggcgtggggc ctcgagcag ggcggccagg	1260
atgccagatg attattctgg agtctgggat cgggtgtgcc ggggaacgga cacggggctg	1320
gactgctcgc ggggtcgttg cacaggggct gagctacca gcgatactgg tgttcgaaat	1380
aagagtgcga ggcaaggac cagacagtgc tggggactgg gattattccg gggactcgca	1440
cgtgaattgg atgccaagga ataacggtga ccaggaaagg cggggaggca ggatggcggt	1500
agagattgac gatggtctca aggacggcgc gcaggtgaag ggggtgttg gcgatggctg	1560
cgcccaggaa caagggtggc cggctctggct gtgcgtgatg gccaggcggt agcataatga	1620
cggaatacag aggaggcgag tgagtggcca gggagctgga gattctgggg tccagggcaa	1680
agataatctg cccccgactc ccagtctctg atgcaaaacc gagtgaaccg ttataccagc	1740

cttgccattt taagaattac ttaagggccg ggcgcggtgg ccactcctg taatcccagc 1800
 actttgggag gccgaggcgg atggatcact tgaagtcagg agttgaccag cctggccaac 1860
 atggtgaaag cctgtctcta ccaaaaatag aaaaattaat cgggcgctat ggcgggtgcc 1920
 ttaatcccag ctactcgggg gggctaaggc aggagaatcg cttgaaccgg ggaggcggag 1980
 gtttcagtga gccgagatcg cgccactgca ctccagcctg ggccagagtg agactccgtc 2040
 tcaaaaaaaaa aaaaaaaaaa aaaaaaaaaa agacttactt aaggtctaag atgaaaagca 2100
 gggcctacgg agtagccacg tccgggacctg gtctggggag aggggaggat agggtcagtg 2160
 acatggaatc ctgacgtggc caaagggtgcc cgggtgccagg agatcatcga cccttggact 2220
 aggatgggag gtcggggaac agaggatagc ccagggtggct tcttggaat cacctttctc 2280
 gggcagggtc caaggcactg ggttgacagt cctaacctgg ttccaccca cccacccct 2340
 ctgccaggtg gggcaggggt ttccggctgg tggagcatgt gctgggacag gacagcatcc 2400
 tcaatcaatc caacagcata ttcggttgca tcttctacac actacagcta ttgttaggtg 2460
 agtggctccg cccctccct gcccgccccg ccccgccct catccccctt ggtcagctca 2520
 gcccactcc atgcaatctt ggtgatccac acagctgaca gccagctagc tgctcatcac 2580
 ggagcgtcct gcgggtgggg atgtggggag gtaactaaca ggagtctttt aattggttta 2640
 agtactgtta gaggtgaag ggccttaaa gacatcctag gtccccagggt tttttgttg 2700
 ttgttgttt gagacagggt ctggctctgt tgccaaagt gaggtctagg atgcccttag 2760
 tgtgactgg cgtgatctca gttcatggca acctctgct ccctgccccaa gggatccctc 2820
 caccttagcc tccaagcag ctggaatcac aggcgtgcac cactatgccc agctaatttt 2880
 tgtttttgtt ttttttgggt agagatgggt tctcgccatg ttgccaggc tggctcaag 2940
 caatctgtct gcctcagcct cccaaagtgc tggggggatt acaggcgtga gctaccatgc 3000
 cccaccaaca cccagtttt gtgaaaaga tgccgaaatt cttttttaag gagaagctga 3060
 gcatgagcta tcttttctct catttagtgc tcagcaggaa aatttgatc tagtcccata 3120
 agaacagaga gaggaaccaa gggagtggaa gacgatggcg cccaggcct tgctgatgcc 3180
 atatgccgga gatgagacta tccattacca ccctcccag caggctccca cgctcccttt 3240
 gagtcaccct tcccagctcc agagaaggca tcttgaggg agggccagca ccatggctct 3300
 ggctgacaca tggttcagac ttggccgatt tatttaagaa attttattgc tcagaacttt 3360
 ccctccctgg gcaatggcaa gagcttcaga gaccagtccc ttggagggga cctgttgaag 3420
 ccttcttttt tttttttttt aagaaataat cttgctctgt tgcccaggct ggagtgcagt 3480
 ggcacaatca tagctcactg taacctggct caagcgatcc tcctgagtag ctaggactat 3540

aggcatgtca ctgcacccag ctaatTTTTT tttttttttt tttttttttt ttgcgacata 3600
gtctcgctct gtcaccaggc tggagtgcag tggcacgac ttggctcact gcaacctctg 3660
cctccccggg tcaagcaatt ttcctgcctc agcctcctga gtagctggga ctacaggcgc 3720
gtgtcaccac gccagctaa tttttgtatt tttagtggag acagggtttc accatgttgg 3780
ctaggatggg ctcaatctct tgacctggg atccatccgc cttggcctcc caaagtgcta 3840
ggattacagg cgtgagtcaa cctcacggg catttttttt ttgagacgaa gtcttgctct 3900
tgctgcccc gctggaatgt ggtggcatga tctcggtca ctgcaacctc cacctcctag 3960
gttcaagcga ttctccacct tagcctcccc agcagctggg attacagggtg cccatcaaca 4020
caccgggcta atttttgtat ttttattaga gatgggggtt tgccatgttg gccaggctgc 4080
tctcgaactc ctaacctcag gtgatccacc cccattggcc tcccaaaata ctgggattac 4140
aggcatgagc caccgtgccc agctgaattt ctaaattttt gatagagatc gggctcttct 4200
atgttgcccc agctggtctt gaactcctag cctaaagcag tcttcccacc tcggcctccc 4260
agagtgtttg gaatacgtgc gtaagccacc acatctgccc tggagcctct tgttttagag 4320
acccttcccc gcagctcctg gcatctaggt agtgcagtga catcatggag tgctcgggag 4380
gtggccagtg cctgaagccc acaccggacc ctcttctgcc ttgcagggtg cctgcggaca 4440
cgctgggcct ctgtcctgat gctgctgagc tccctgggtg ctctcgctgg ttctgtctac 4500
ctggcctgga tctgttctt cgtgctctat gatttctgca ttgtttgtat caccacctat 4560
gctatcaacg tgagcctgat gtggctcagt ttccggaagg tccaagaacc ccagggcaag 4620
gctaagaggc actgagccct caacccaagc caggctgacc tcatctgctt tgctttggca 4680
tgtgagcctt gcctaagggg gcatatctgg gtccctagaa ggccctagat gtggggcttc 4740
tagattaccc cctcctcctg ccataccac acatgacaat ggaccaaag tgccacacgc 4800
tcgctctttt ttacaccag tgctctgac tctgtccca tgggctgggc tccaaagctc 4860
ttccattgc ccagggaggg aaggttctga gcaataaagt ttcttagatc aatcagccaa 4920
gtctgaacca tgtgtctgcc atggactgtg gtgctgggccc tccctcggtg ttgccttctc 4980
tgagctggg aagggtgagt cagagggaga gtggagggcc tgctgggaag ggtggttatg 5040
ggtagtctca tctccagtgt gtggagtcag caaggcctgg ggcaccattg gccccaccc 5100
ccaggaaaca ggctggcagc tcgctcctgc tgcccacagg agccaggcct cctctcctgg 5160
gaaggctgag cacacacctg gaagggcagg ctgcccttct ggttctgtaa atgcttgctg 5220
ggaagtctt ccttgagttt aactttaacc cctccagttg ccttatcgac cattccaagc 5280
cagtattggg agccttgag ggtcagggcc aggttgtgaa ggtttttgtt ttgcctatta 5340
tgccctgacc acttacctac atgccaagca ctgtttaaga acttgtgttg gcagggtgca 5400

gtggctcaca cctgtaatcc ctgtactttg ggaggccaag gcaggaggat cacttgaggc 5460
 caggagtcc agaccagcct gggcaaaata gtgagacccc tgtctctaca aaaaaaaaaa 5520
 aaaaaaaaaa ttagccaggc atgggtggtg atgtacctat agtcccaact aatcggaag 5580
 ctggcgggaa gactgcttga gccagaagg ttgaggctgc agtgagccat gatcactgca 5640
 ctccagcctg agcaacagag caagaccgtc tccaaaaaaa aacaaaaaac aaaaaaaac 5700
 ttgtgttaac gtgttaaact cgtttaatct ttacagtgat ttatgagggtg ggtactatta 5760
 ttatccctat cttgatgata gggacagagt ggctaattag tatgcctgag atcacacagc 5820
 tactgcagga ggctctcagg atttgaatcc acctgggtcca tctgggtcca gcactatat 5880
 gctttttttt ttgttggttt gtttttgaga cggac 5915

<210> 15

<211> 5915

<212> DNA

<213> Homo sapiens

<400> 15

caccatcaga tgggacgtct gtgaaggaga gacctcatct ggcccacagc ttggaaagga 60
 gagactgact gttgagttga tgcaagctca ggtgttgcca ggcgggccc atgatagtag 120
 agaggttagg atactgtcaa ggggtgtgtg ggccaaagga gtggttctgt gaatgtatgg 180
 gagaaagga gaccgaccac caggaagcac tggtaggca ggaccggga ggatgggagg 240
 ctgcagcccc aatggtgcct gaaatagttt caggggaaat gcttggttcc cgaatcggat 300
 cgccgtattc gctggatccc ctgatccgct ggtctctagg tcccgatgc tgcaattctt 360
 acaacaggac ttggcatagg gtaagcgcaa atgctgttaa ccacactaac acactttttt 420
 tttctttttt tttttttgag acagagtctc actctgtcgg cctggctgga gtgcagtggc 480
 acgatctcgg ctactgcaa cctccggctc cccggctcaa gcaattctcc tgccctagcc 540
 tcccgagtag ctgggattac agacatgtgc caccacgccc ggctaatttt tgtattttta 600
 gttgagatgg ggtttcacca tgttggcgag gctggtcttg aactcctgac ctccaggaat 660
 ccgccagcct cggcctccca aagtgtggg attacaagcg tgagccaccg tgcccggcca 720
 acagttttta aatctgtgga gacttcattt cccttgatgc cttgcagccg cgccgactac 780
 aactcccatc atgcctggca gccgctgggg ccgcgattcc gcacgtccct taccgcttc 840
 actagtcccc gcattcttcg ctgttttcct aactcgcccc cttgactagc gccctggaac 900
 agccatttgg gtcgtggagt gcgagcacgg ccggccaatc gccgagtcag agggccagga 960
 ggggcgcgcc cattcgccgc ccggcccctg ctccgtggct ggttttctcc gcgggcgcct 1020
 cgggcggaac ctggagataa tgggcagcac ctgggggagc cctggctggg tgccgctcgc 1080

tctttgcctg acgggcttag tgctctcgct ctacgcgctg cacgtgaagg cggcgcgcg 1140
ccgggaccgg gattaccgcg cgctctgcga cgtgggcacc gccatcagct gttcgcgcg 1200
cttctcctcc aggtgtgcac gggagtggga ggcgtggggc ctcgagcag ggcggccagg 1260
atgccagatg attattctgg agtctgggat cgggtgtgcc ggggaacgga cacggggctg 1320
gactgctcg cgggctcgtt cacaggggct gagctacca gcgatactgg tgttcgaaat 1380
aagagtgcga ggcaaggac cagacagtgc tggggactgg gattattccg gggactcgca 1440
cgtgaattgg atgccaagga ataacggtga ccaggaaagg cggggaggca ggatggcgg 1500
agagattgac gatggtctca aggacggcg gcaggtgaag gggggtgtt gcgatggctg 1560
cgcccaggaa caaggtggcc cggctctggct gtgcgtgatg gccaggcgtt agcataatga 1620
cggaatacag aggaggcgag tgagtggcca gggagctgga gattctgggg tccagggcaa 1680
agataatctg ccccgactc ccagtctctg atgcaaaacc gagtgaaccg ttataccagc 1740
cttgccattt taagaattac ttaagggccg ggcgcggtgg ccactcctg taatcccagc 1800
actttgggag gccgagcgag atggatcact tgaagtcagg agttgaccag cctggccaac 1860
atggtgaaag cctgtctcta ccaaaaatag aaaaattaat cgggcgctat ggcgggtgcc 1920
ttaatcccag ctactcgggg gggctaaggc aggagaatcg cttgaaccgg ggaggcggag 1980
gtttcagtga gccgagatcg cgccactgca ctccagcctg ggccagagtg agactccgtc 2040
tcaaaaaaaaa aaaaaaaaaa aaaaaaaag agacttactt aaggtctaag atgaaaagca 2100
gggcctacgg agtagccacg tccgggctg gtctggggag aggggaggat agggtcagtg 2160
acatggaatc ctgacgtggc caaaggtgcc cgggtgccagg agatcatcga cccttgact 2220
aggatgggag gtcggggaac agaggatagc ccaggtggct tcttggaat cacctttctc 2280
gggcagggtc caaggcactg ggttgacagt cctaacctgg ttccaccca cccaccct 2340
ctgccaggtg ggcaggggt ttcgggctgg tggagcatgt gctgggacag gacagcatcc 2400
tcaatcaatc caacagcata ttcggttgca tcttctacac actacagcta ttgttaggtg 2460
agtggctccg cccctccct gcccgccccg ccccgccct catccccctt ggtcagctca 2520
gccccactcc atgcaatctt ggtgatccac acagctgaca gccagctagc tgctcatcac 2580
ggagcgtcct gcgggtgggg atgtggggag gtaactaaca ggagtctttt aattggttta 2640
agtactgtta gaggctgaag ggccctaaa gacatcctag gtccccagg tttttgtttg 2700
ttgttgtttt gagacagggt ctggctctgt tgccaaagt gaggtctagg atgcccttag 2760
tgtgcactgg cgtgatctca gttcatggca acctctgcct ccctgccccaa gggatccctc 2820
caccttagcc tccaagcag ctggaatcac aggcgtgcac cactatgccc agctaatttt 2880

tgtttttgggt	tttttttgggt	agagatgggtg	tctcgccatg	ttgcccaggc	tggtctcaag	2940
caatctgtct	gcctcagcct	cccaaagtgc	tggtgggatt	acaggcgtga	gctaccatgc	3000
cccaccaaca	ccccagtttt	gtggaaaaga	tgccgaaatt	cctttttaag	gagaagctga	3060
gcatgagcta	tcttttgtct	catttagtgc	tcagcaggaa	aatttgtatc	tagtcccata	3120
agaacagaga	gaggaacca	gggagtggaa	gacgatggcg	ccccaggcct	tgctgatgcc	3180
atatgccgga	gatgagacta	tccattacca	cccttcccag	caggctccca	cgctcccttt	3240
gagtcaccct	tcccagctcc	agagaaggca	tactgaggg	aggcccagca	ccatggctct	3300
ggctgacaca	tggttcagac	ttggccgatt	tatttaagaa	attttattgc	tcagaacttt	3360
ccctccctgg	gcaatggcaa	gagcttcaga	gaccagtccc	ttggagggga	cctgttgaag	3420
ccttcttttt	tttttttttt	aagaaataat	cttgctctgt	tgcccaggct	ggagtgcagt	3480
ggcacaatca	tagctcactg	taacctggct	caagcgatcc	tcctgagtag	ctaggactat	3540
aggcatgtca	ctgcacccag	ctaatttttt	tttttttttt	tttttttttt	ttgcgacata	3600
gtctcgctct	gtcaccaggc	tggtgtgcag	tggtcacgatc	ttggctcact	gcaacctctg	3660
cctcccgggt	tcaagcaatt	ttcctgcctc	agcctcctga	gtagctggga	ctacaggcgc	3720
gtgtcaccac	gcccagctaa	tttttgtatt	tttagtggag	acagggtttc	accatgttgg	3780
ctaggatggt	ctcaatctct	tgacctgggtg	atccatccgc	cttggcctcc	caaagtgcta	3840
ggattacagg	cgtgagtcaa	cctcaccggg	catttttttt	ttgagacgaa	gtcttgcctc	3900
tgctgcccaa	gctggaatgt	ggtggcatga	tctcggtcca	ctgcaacctc	cacctcctag	3960
gttcaagcga	ttctccacct	tagcctcccc	agcagctggg	attacagggtg	cccatcaaca	4020
cacccggcta	atttttgtat	ttttattaga	gatgggggtt	tgccatgttg	gccaggctgc	4080
tctcgaactc	ctaacctcag	gtgatccacc	cccattggcc	tcccaaaata	ctgggattac	4140
aggcatgagc	caccgtgccc	agctgaattt	ctaaattttt	gatagagatc	gggtctttct	4200
atgttgccca	agctggctct	gaactcctag	cctaaagcag	tcttcccacc	tcggcctccc	4260
agagtgtttg	gaatacgtgc	gtaagccacc	acatctgccc	tgtagcctct	tgtttttagag	4320
acccttccca	gcagctcctg	gcatctaggt	agtgcagtga	catcatggag	tggtcgggag	4380
gtggccagtg	cctgaagccc	acaccggacc	ctcttctgcc	ttgcagggtg	cctgcggaca	4440
cgctgggcct	ctgtcctgat	gctgctgagc	tccctgggtg	ctctcgctgg	ttctgtctac	4500
ctggcctgga	tctgttctt	cgtgctctat	gatttctgca	ttgtttgtat	caccacctat	4560
gctatcaacg	tgagcctgat	gtggctcagt	ttccggaagg	tccaagaacc	ccagggcaag	4620
gctaagaggc	actgagccct	caaccaagc	caggctgacc	tcatctgctt	tgctttggca	4680
tgtgagcctt	gcctaagggg	gcatactctg	gtccctagaa	ggccctagat	gtggggcttc	4740

```

tagattaccc cctcctcctg ccataccgc acatgacaat ggaccaaagtg tgccacacgc 4800
tcgctctttt ttacaccagc tgcctctgac tctgtcccca tgggctgggc tccaaagctc 4860
tttccattgc ccagggaggg aagggtctga gcaataaagt ttcttagatc aatcagccaa 4920
gtctgaacca tgtgtctgcc atggactgtg gtgctggggc tccctcggtg ttgccttctc 4980
tgagctggg aagggtgagt cagagggaga gtggagggcc tgctgggaag ggtggttatg 5040
ggtagtctca tctccagtgt gtggagtcag caaggcctgg ggcaccattg gccccaccc 5100
ccaggaaca ggctggcagc tcgctcctgc tgcccacagg agccaggcct cctctcctgg 5160
gaaggctgag cacacacctg gaagggcagg ctgcccttct ggttctgtaa atgcttctg 5220
ggaagttctt ccttgagttt aactttaacc cctccagtgt ccttatcgac cattccaagc 5280
cagtattggt agccttgagg ggtcagggcc aggttggtga ggtttttgtt ttgcctatta 5340
tgccctgacc acttacctac atgccaagca ctgtttaaga acttgtgttg gcagggtgca 5400
gtggctcaca cctgtaatcc ctgtactttg ggaggccaag gcaggaggat cacttgaggc 5460
caggagtcc agaccagcct gggcaaaata gtgagacccc tgtctctaca aaaaaaaaaa 5520
aaaaaaaaa ttagccaggc atgggtggtg atgtacctat agtcccaact aatcggaag 5580
ctggcgggaa gactgcttga gcccagaagg ttgaggctgc agtgagccat gatcactgca 5640
ctccagcctg agcaacagag caagaccgtc tccaaaaaaa aacaaaaaac aaaaaaaac 5700
ttgtgttaac gtgttaaaact cgtttaatct ttacagtgat ttatgaggtg ggtactatta 5760
ttatccctat cttgatgata gggacagagt ggctaattag tatgcctgag atcacacagc 5820
tactgcagga ggctctcagg atttgaatcc acctggtcca tctggctcca gcatttatat 5880
gctttttttt ttgttggttt gtttttgaga cggac 5915

```

<210> 16

<211> 5915

<212> DNA

<213> Homo sapiens

<400> 16

```

caccatcaga tgggacgtct gtgaaggaga gacctcatct ggcccacagc ttggaaagga 60
gagactgact gttgagttga tgcaagctca ggtgttgcca ggcggcgccc atgatagtag 120
agaggtagg atactgtcaa ggggtgtgtg ggccaaagga gtggttctgt gaatgtatgg 180
gagaaaggga gaccgaccac caggaagcac tggtaggca ggaccggga ggatgggagg 240
ctgcagccc aatgggtgct gaaatagttt caggggaaat gcttggttcc cgaatcggat 300
cgccgtattc gctggatccc ctgatccgtt ggtctctagg tcccggatgc tgcaattctt 360
acaacaggac ttggcatagg gtaagcgcaa atgctgttaa ccacactaac acactttttt 420

```

ttttcttttt ttttttgag acagagtctc actctgtcgg cctggctgga gtgcagtggc	480
acgatctcgg ctactgcaa cctccggctc cccggctcaa gcaattctcc tgcctcagcc	540
tcccagtag ctgggattac aggcattgtc caccacgccc ggctaatttt tgtattttta	600
gttgagatgg ggtttcacca tgttggcgag gctggtcttg aactcctgac ctgaggtaat	660
ccgccagcct cggcctccca aagtgtggg attacaagcg tgagccaccg tgcccggcca	720
acagttttta aatctgtgga gacttcattt cccttgatgc cttgcagccg cgccgactac	780
aactcccatc atgcctggca gccgctggg cgcgattcc gcacgtccct taccgcttc	840
actagtcctg gcattcttcg ctgttttctt aactcgcccg cttgactagc gccctggaac	900
agccatttgg gtctgtggagt gcgagcacgg ccggccaatc gccgagtcag agggccagga	960
ggggcgcggc cattcgccgc cgggccctg ctccgtggct ggttttctcc gcgggcgcct	1020
cgggcggaac ctggagataa tgggcagcac ctgggggagc cctggctggg tgcggtcgc	1080
tctttgcctg acgggcttag tgctctcgt ctacgcgtg cactgaagg cggcgcgcg	1140
ccgggaccgg gattaccgcg cgctctcga cgtgggcacc gccatcagct gttcgcgct	1200
cttctcctcc aggtgtgcac gggagtggga ggcgtggggc ctcgagcag ggcggccagg	1260
atgccagatg attattctgg agtctgggat cgggtgtgcc ggggaacgga cacggggctg	1320
gactgctcgc ggggtcgttg cacaggggct gagctacca gcgatactgg tgttcgaaat	1380
aagagtgcga ggcaaggac cagacagtgc tggggactgg gattattccg gggactcgca	1440
cgtgaattgg atgccaagga ataacggtga ccaggaaagg cggggaggca ggatggcgg	1500
agagattgac gatggtctca aggacggcg gcaggtgaag gggggtgtg gcgatggctg	1560
cggccaggaa caaggtggcc cggctctggt gtgcgtgatg gccaggcgtt agcataatga	1620
cggataacag aggagggcag tgagtggcca gggagctgga gattctggg tccagggcaa	1680
agataatctg ccccgactc ccagtctctg atgcaaaacc gactgaaccg ttataaccagc	1740
cttgccattt taagaattac ttaagggcgg ggcgcggtgg ccactcctg taatcccagc	1800
actttgggag gccgaggcgg atggatcact tgaagtcagg agttgaccag cctggccaac	1860
atggtgaaag cctgtctcta ccaaaaatag aaaaattaat cgggcgctat ggcggtgcc	1920
ttaatcccag ctactcgggg gggctaaggc aggagaatcg cttgaaccgg ggaggcggag	1980
gtttcagtga gccgagatcg cgccactgca ctccagcctg ggccagagtg agactccgtc	2040
tcaaaaaaaaa aaaaaaaaaa aaaaaaaaaa agacttactt aaggcttaag atgaaaagca	2100
gggcctacgg agtagccacg tccgggcctg gtctggggag aggggaggat agggtcagt	2160
acatggaatc ctgacgtggc caaaggtgcc cggtgccagg agatcatcga cccttggaact	2220

aggatgggag gtcggggaac agaggatagc ccagggtggct tcttggaat caccctttctc 2280
 gggcagggtc caaggcactg ggttgacagt cctaacctgg ttccaccca cccacccct 2340
 ctgccagggtg gggcaggggt ttcgggctgg tggagcatgt gctgggacag gacagcatcc 2400
 tcaatcaatc caacagcata ttcgggttgca tcttctacac actacagcta ttgttaggtg 2460
 agtggtccg cccctccct gcccgcccg ccccgccct catccccctt ggtcagctca 2520
 gcccactcc atgcaatctt ggtgatccac acagctgaca gccagctagc tgctcatcac 2580
 ggagcgtcct gcgggtggg atgtggggag gtaactaaca ggagtctttt aattggttta 2640
 agtactgtta gaggtgaag ggccttaaa gacatcctag gtccccagggt tttttgtttg 2700
 ttgttgtttt gagacagggt ctggctctgt tgccaaagt gaggtctagg atgcccttag 2760
 tgtgactgg cgtgatctca gttcatggca acctctgect ccctgccccaa gggatccctc 2820
 cacccttagcc tccaagcag ctggaatcac aggcgtgcac cactatgccc agctaatttt 2880
 tgtttttgtt ttttttttgt agagatgggt tctcgccatg ttgcccaggc tggctcctcaag 2940
 caatctgtct gcctcagcct cccaaagtgc tggggggatt acaggcgtga gctaccatgc 3000
 cccaccaaca cccagtttt gtggaaaaga tgccgaaatt cttttttaag gagaagctga 3060
 gcatgagcta tcttttgtct catttagtgc tcagcaggaa aatttgtatc tagtcccata 3120
 agaacagaga gaggaacca gggagtggaa gacgatggcg cccaggcct tgctgatgcc 3180
 atatgccgga gatgagacta tccattacca cccttcccag caggctccca cgctcccttt 3240
 gagtcaccct tcccagctcc agagaaggca tctactgagg aggccagca ccatggtcct 3300
 ggctgacaca tgggtcagac ttggccgatt tattaagaa attttattgc tcagaacttt 3360
 ccctccctgg gcaatggcaa gagcttcaga gaccagtccc ttggagggga cctgttgaag 3420
 ccttcttttt tttttttttt aagaaataat cttgctctgt tgcccaggct ggagtgcagt 3480
 ggcacaatca tagtctactg taacctggct caagcgatcc tctgagtag ctaggactat 3540
 aggcattgtca ctgacccag ctaatttttt tttttttttt tttttttttt ttgcgacata 3600
 gtctcgtct gtcaccaggc tggagtgcag tggcacgac ttggctcact gcaacctctg 3660
 cctcccggt tcaagcaatt ttcctgcctc agcctcctga gtagctggga ctacaggcgc 3720
 gtgtcaccac gccagctaa tttttgtatt tttagtggag acagggtttc accatgttgg 3780
 ctaggatggt ctcaatctct tgacctggtg atccatccgc cttggcctcc caaagtgcta 3840
 ggattacagg cgtgagtcaa cctcaccggg catttttttt ttgagacgaa gtcttgctct 3900
 tgctgcccc gctggaatgt ggtggcatga tctcggtca ctgcaacctc cactcctag 3960
 gttcaagcga ttctccacct tagcctcccc agcagctggg attacagggtg cccatcaaca 4020
 caccggcta atttttgtat ttttattaga gatgggggtt tgccatgttg gccaggctgc 4080

tctcgaactc ctaacctcag gtgatccacc cccattggcc tcccaaaata ctgggattac	4140
aggcatgagc caccgtgccc agctgaattt ctaaattttt gatagagatc gggctctttct	4200
atgttgccca agctggtctt gaactcctag cctaaagcag tcttcccacc tcggcctccc	4260
agagtgtttg gaatacgtgc gtaagccacc acatctgccc tggagcctct tgttttagag	4320
acccttccca gcagctcctg gcatctaggt agtgacgtga catcatggag tgttcgggag	4380
gtggccagtg cctgaagccc acaccggacc ctcttctgcc ttgcaggttg cctgcggaca	4440
cgctgggcct ctgtcctgat gctgctgagc tccctgggtg ctctcgctgg ttctgtctac	4500
ttggcctgga tctgttctt cgtgctctat gatttctgca ttgtttgtat caccacctat	4560
gctatcaacg tgagcctgat gtggctcagt ttccggaagg tccaagaacc ccagggcaag	4620
gctaagaggc actgagccct caaccaagc caggctgacc tcatctgctt tgctttggca	4680
tgtgagcctt gcctaagggg gcatatctgg gtccctagaa ggccctagat gtggggcttc	4740
tagattaccc cctcctcctg ccataccgc acatgacaat ggaccaaagtg tgccacacgc	4800
tcgctctttt ttacaccag tgcctctgac tctgtcccca tgggctgggc tccaaagctc	4860
tttccattgc ccagggaggg aaggttctga gcaataaagt ttcttagatc aatcagccaa	4920
gtctgaacca tgtgtctgcc atggactgtg gtgctgggccc tccctcggtg ttgccttctc	4980
tggagctggg aagggtgagt cagagggaga gtggagggcc tgctgggaag ggtggttatg	5040
ggtagtctca tctccagtgt gtggagtcag caaggcctgg ggcaccattg gccccaccc	5100
ccaggaaaca ggctggcagc tcgctcctgc tgcccacagg agccaggcct cctctcctgg	5160
gaaggctgag cacacacctg gaagggcagg ctgcccttct ggttctgtaa atgcttgctg	5220
ggaagtctct ccttgagttt aactttaacc cctccagttg ccttatcgac cattccaagc	5280
cagtattggt agccttgag ggtcagggcc aggttgtaga ggtttttgtt ttgcctatta	5340
tgccctgacc acttacctac atgccaaagca ctgtttaaga acttggtgtg gcaggggtgca	5400
gtggctcaca cctgtaatcc ctgtactttg ggaggccaag gcaggaggat cacttgaggc	5460
caggagttcc agaccagcct gggcaaaata gtgagacccc tgtctctaca aaaaaaaaaa	5520
aaaaaaaaaa ttagccaggc atgggtggtg atgtacctat agtcccaact aatcgggaag	5580
ctggcgggaa gactgcttga gcccagaagg ttgaggctgc agtgagccat gatcactgca	5640
ctccagcctg agcaacagag caagaccgtc tccaaaaaaa aacaaaaaac aaaaaaaac	5700
ttgtgttaac gtgttaaact cgtttaatct ttacagtgat ttatgaggtg ggtactatta	5760
ttatccctat cttgatgata gggacagagt ggctaattag tatgcctgag atcacacagc	5820
tactgcagga ggctctcagg atttgaatcc acctggtcca tctggctcca gcacttatat	5880

gctttttttt ttgttggttt gtttttgaga cggac 5915

<210> 17
<211> 15
<212> DNA
<213> Artificial

<220>
<223> vk2581 G>C VIC probe sequence

<400> 17
tcatcacgga gcgtc 15

<210> 18
<211> 15
<212> DNA
<213> Artificial

<220>
<223> vk2581 G>C FAM probe sequence

<400> 18
tcatcacgga gcgtc 15

<210> 19
<211> 20
<212> DNA
<213> Artificial

<220>
<223> PCR primer

<400> 19
ggtgatccac acagctgaca 20

<210> 20
<211> 23
<212> DNA
<213> Artificial

<220>
<223> PCR primer

<400> 20
octgttagtt acctccccac atc 23

<210> 21
<211> 15
<212> DNA
<213> Artificial

<220>
<223> vk3294 T>C VIC probe sequence

<400> 21
ccaggaccat ggtgc 15

<210> 22
<211> 15
<212> DNA
<213> Artificial

<220>
<223> vk3294 T>C FAM probe sequence

<400> 22
ccaggaccgt ggtgc 15

<210> 23
<211> 20
<212> DNA
<213> Artificial

<220>
<223> PCR primer

<400> 23
gctccagaga aggcatact 20

<210> 24
<211> 22
<212> DNA
<213> Artificial

<220>
<223> PCR primer

<400> 24
gccaaagtctg aacctgtgt ca 22

<210> 25
<211> 15
<212> DNA
<213> Artificial

<220>
<223> vk4769 G>A VIC probe sequence

<400> 25
ataccgcac atgac 15

<210> 26
<211> 16
<212> DNA
<213> Artificial

<220>
<223> vk4769 G>A FAM probe sequence

<400> 26
cataccaca catgac 16

<210> 27
<211> 22
<212> DNA
<213> Artificial

<220>
<223> PCR primer

<400> 27
gtccctagaa ggccctagat gt

22

<210> 28
<211> 21
<212> DNA
<213> Artificial

<220>
<223> PCR primer

<400> 28
gtgtggcaca tttggtccat t

21

<210> 29
<211> 19
<212> DNA
<213> Artificial

<220>
<223> PCR primer

<400> 29
ccaatcgccg agtcagagg

19

<210> 30
<211> 20
<212> DNA
<213> Artificial

<220>
<223> PCR primer

<400> 30
cccagtcgcc agcactgtct

20

<210> 31
<211> 20
<212> DNA
<213> Artificial

<220>
<223> PCR primer

<400> 31
aggggaggat agggtcagtg

20

<210> 32
<211> 21
<212> DNA
<213> Artificial

<220>
<223> PCR primer

<400> 32
cctgtagtt acctccccac a 21

<210> 33
<211> 20
<212> DNA
<213> Artificial

<220>
<223> PCR primer

<400> 33
atacgtgcgt aagccaccac 20

<210> 34
<211> 20
<212> DNA
<213> Artificial

<220>
<223> PCR primer

<400> 34
accagatat gcccccttag 20